The Dissertation Committee for Raghu G. Raj
certifies that this is the approved version of the following dissertation:

Optimal Visual Search Strategies using Natural Scene Statistics

Committee:

_____
Alan C. Bovik, Supervisor

_____
Gustavo de Veciana

_____
Wilson S. Geisler

_____
Joydeep Ghosh

_____
John E. Gilbert

_____
Ross Baldick

Optimal Visual Search Strategies using Natural Scene Statistics

by

Raghu G. Raj, B.S.E.E.; B.S.C.S.; M.S.E.E.

DISSERTATION

Presented to the Faculty of the Graduate School
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
of the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
December 2007

Om Sri Gurubhyo Namaha

Acknowledgements

I would firstly like to express my deep gratitude to my wonderful advisor Prof. Al Bovik because of whose patience I was able to persevere with this dissertation. He has extended his support and encouragement to me through the highs and lows, through the thick and thin. He has taught me by his personal example how to lead a balanced life and to have the correct outlook towards research. It was Prof. Bovik who first initiated me into the vast field of Visual Search and inspired into me its deeper significance in computational vision and image processing.

I would also like to tremendously thank Prof. Wilson Geisler with whom I had the great opportunity of working with during my thesis research. It was by interaction with Prof. Geisler that I came to realize the deeper significance of Natural Scene Statistics modeling in computational vision—something that has guided my research work ever since.

It was also a great privilege to interact with Prof. Gustavo de Veciana. In spite of his exceedingly busy schedule, Prof. Gustavo somehow always found time to talk to me about my work and to give valuable words of advice. I also very much thank Prof. Joydeep Ghosh, my remaining committee members (Prof. Ross Baldick and Prof. John Gilbert), and also Prof. Lawrence Cormack for encouraging me in my work in every possible way.

To me all these gentlemen are model scientists whose dedication to their field of science is truly inspiring and whose character and behavior truly exemplary.

I am also very much indebted to the many friends in LIVE for their support and friendship. I am especially indebted to Umesh for his untiring and selfless help and

encouragement not only to me but to the entire LIVE lab. He is the one person we all look to for help and support in every way. I am also indebted to Sumohana for his friendship and patience with me; and also to the other present and former LIVE members including Mehul, Kalpana, Hamid, Farooq, Yang, Abtine, Junsoo, James, Sina, and all of the recent members of LIVE etc. for their support. Beyond LIVE too I have many friends of whom I would like to single out Amit, Mrinal, Vishal, Akito, Johan and Varuna for their constant friendship and support throughout the years.

Of course I wouldn't have achieved anything in life without the love and encouragement of my family. I have no words to adequately describe how much they have helped to shaped my thinking and the course of my life.

Apart from this I also thank all other people whom I may have carelessly neglected to acknowledge. But since we are all part of the One Divine who lives through and permeates all, Dedication of all works to Him is the most fitting way to summarize the true substance behind all acknowledgments.

Optimal Visual Search Strategies using Natural Scene Statistics

Publication No._____

Raghu G. Raj, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Alan C. Bovik

I present theoretical foundations and perform computational studies on optimal search strategies in natural scenes performed by foveated artificial vision systems, based on novel characterizations of Natural Scene Statistics (NSS).

I first develop relevant theoretical bounds on the processing of foveated—more generally LSV-filtered (Linear Scale Variant)—signals, which provide a rigorous basis to linear post-processing operations performed on foveated images. The major contribution of this dissertation, however, lies in the discovery and elucidation of two major statistical characterizations of natural scenes and their subsequent deployment for devising optimal fixation strategies.

The first is a novel characterization of the contrast statistics of natural scenes, parameterized by the eccentricity at which each contrast level is measured across the LSV-filtered image. This formulation of contrast statistics finds natural application in devising fixation patterns that optimally extract contrast information from the image. I further demonstrate that the resulting fixation patterns are nearly optimal in the sense of minimizing the global MSE of the LSV-filtered image.

The second is the characterization of the non-stationary structure of natural images via the development of the concept of non-stationarity indices that measure the extent of

non-stationarity across the image. The theoretical motivation of our approach lies in a novel characterization of image patch statistics I developed, called Multilinear Independent Component Analysis (MICA), wherein the statistical interactions between the pseudo-independent components are captured via a multilinear expansion of the joint probability density being modeled. This modeling technique enables the derivation of a theoretical measure of non-stationary in natural scenes that subsequently motivates computationally efficient non-stationarity indices—a variant of which is then deployed to furnish optimal texture-based fixations natural images. The fixation patterns generated by our information-theoretic approaches are quantitatively shown to match very well with human fixation patterns and offer considerable explanatory and predictive power over previously well-known fixation strategies.

These results point the way towards a unified information-theoretic understanding of low-level fixation processes; and further demonstrate the importance of incorporating low-level visual information into visual search strategies—thereby providing a foundation upon which high-level visual information relating to scene context and object structures can be incorporated.

Table of Contents

Acknowledgement **v**

Abstract **vii**

List of Tables **xii**

List of Figures **xiii**

Chapter 1. Introduction and Background **1**

1.1 Motivation……1

1.2 A Role for Information Theory in Low-level Fixation Selection……...3

1.3 Related Research……..4

1.3.1 Foveation……..4

1.3.2 Natural Scene Statistics Models……...5

1.3.3 High-level Visual Search……9

1.3.4 Low-level Visual Search……11

1.4 Dissertation Overview……...14

Chapter 2. Approximating Filtered Scale-Variant Signals **19**

2.1 Introduction……19

2.2 Quasi-invariant Filtering of Scale-Variant Signals……21

2.3 Discrete Formulation……..34

2.4 Simulation Results……36

2.5 Conclusions……38

Chapter 3. Contrast Statistics of Natural Images: Fixation Selection by Minimization of Contrast Entropy **43**

3.1 Introduction……....43
3.2 Methods and Results……..45
3.2.1 Contrast Statistics……...45
3.2.2 Fixation Selection……..50
3.2.2.1 Contrast Entropy Minimization Algorithm…….51
3.2.2.2 Performance of the CEM Algorithm……...52
3.3 Discussion……..57

Chapter 4. MICA: A Multilinear ICA Decomposition for Natural Image Modeling 65
4.1 Introduction……65
4.2 The Multilinear ICA Model…….…..67
4.3 Simulation Results…….77
4.4 Discussion…….….82

Chapter 5. Non-stationarity Measurement in Natural Images 95
5.1 Introduction……....95
5.2 On Non-stationarity Measurement……99
5.3 Simulation Results……...…117
5.4 Discussion……119

Chapter 6. Texture-Contrast Based Fixation Selection in Natural Images 129
6.1 Introduction…….….129
6.2 Contrast-based Fixations……..134
6.3 Texture-based Fixations……...…138
6.4 Combining Texture and Contrast Fixation Features……145

6.5 Discussion……....153

Chapter 7. Contributions and Future Work **172**

7.1 Contributions……...….172

7.2 Future Work…….174

Appendix A **Proofs of Chapter 2** **179**

Appendix B **Proofs of Chapter 3** **188**

Appendix C **Proofs of Chapter 4** **192**

Appendix D **Proofs of Chapter 5** **193**

Bibliography **195**

Vita **213**

List of Tables

4.1 Relative improvement with respect to classic ICA when using the complete basis MICA model is shown for the various textures.....88

4.2 Relative improvement with respect to classic ICA when using the under-complete MICA model.....88

4.3 Relative improvement with respect to classic ICA when using MICA for contrast images.....88

4.4 Relative improvement with respect to classic ICA when using MICA for densely sampled texture regions.....89

6.1-A Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{ave} and single-foveal width Gaussian interpolation.....163

6.1-B Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{harmonic} and single-foveal width Gaussian interpolation.....163

6.2-A Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{ave} and twice-foveal width Gaussian interpolation.....164

6.2-B Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{harmonic} and twice-foveal width Gaussian interpolation.....164

List of Figures

2.1 Foveated Lena Image with four annular regions……...…39

2.2 Defoveated Lena Image corresponding to Fig. 2.1……39

2.3 Quasi-invariant approximation to the defoveated Lena image…...……...…39

2.4 Graded foveated Lena image…….39

2.5 Defoveation of graded foveated Lena image…….……40

2.6 Quasi-invariant approximation of graded defoveated Lena image……....40

2.7 Lena image foveated over four annular regions corrupted by AWGN noise……40

2.8 Defoveated version of graded foveated noisy Lena image in Fig. 2.7……..…40

2.9 MMSE (Wiener) defoveated version of graded foveated noisy image in Fig. 2.7……...41

2.10 Plot of theoretical and actual zero-crossing rates averaged over 100 1-D Gaussian noise signals filtered by scale- variant linear Gaussian filters…...……41

2.11 Left: Scale-variant LoG-filtered Lena image. Right: ZCs computed from Left Image……..……42

2.12 Plot of theoretical and actual zero-crossing rates averaged over 100 radial directions on image filtered by scale-variant linear Gaussian filter. (Legend: +:Theoretical ZC Rate, x: Actual ZC Rate)……...42

3.1 Probability distributions of local rms contrast for various levels of blur based on the human contrast sensitivity function at different retinal eccentricities. These distributions were obtained by randomly sampling small patches from 300 calibrated natural images……..…..61

3.2 These plots show examples of the conditional probability distributions of local rms contrast in unblurred images, given the local rms contrast in the blurred versions of the images (columns) and given the retinal eccentricity (rows). The solid symbols are empirical histograms computed from 300 natural images that contained no man-made objects. The smooth curves are the best-fitting skewed Gaussian distribution (a Gaussian with different standard deviations above and below the mode)…….……61

3.3 Modes and average standard deviations of the conditional probability densities are plotted as a function of blurred image contrast and retinal eccentricity. The average standard deviation is the average of the two standard deviation parameters in the skewed Gaussian distribution. See Fig. 3.2 for examples of the conditional densities and fits of the skewed Gaussian distribution. The curves are best fitting straight lines through the origin……..…62

3.4 Slopes of the linear functions in Fig. 3.3. A, Slope of the contrast versus mode plot as a function of retinal eccentricity. B, Slope of the contrast versus average standard deviation plot as a function of retinal eccentricity. The curves show the predictions of the linear model: $\hat{c} = k\varepsilon c + c$ and $\bar{\sigma} = k\varepsilon c$, where k=0.105…....62

3.5 Images used to test a fixation selection algorithm based on the principle of minimizing contrast entropy……..63

3.6 Fixation points selected by the principle of minimizing total contrast entropy (contrast uncertainty), using the average local contrast statistics of natural images. A, Sequence of nine fixations (eight saccades) for a distant image containing sky, ground, and trees. B, Relative contrast entropy as a function of fixation number for the image in A (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). C, Sequence of nine fixations (eight saccades) for a close-up image containing foliage. D, Same type of plot shown in B……..63

3.7 Average fixation selection performance for the 16 test images in Fig. 5. A, Relative contrast entropy as a function of fixation number (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). B, Ratio of the optimal contrast entropy that could be obtained to the contrast entropy that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles). C, Relative mean squared error (MSE) between the original (unblurred) image and the image reconstructed from the fixations up to and including the fixation number given on the horizontal axis: CEM algorithm (solid circles), optimal (open circles). D, Ratio of optimal MSE

that could be obtained to the MSE that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles)…….64

4.1 Non-linear system model of the multilinear structure of source statistics derived from natural scene models……...84

4.2 (*a*) Gravel (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The high-kurtosis heuristic $\beta_{high-kurt}$ was used…….……84

4.3 (*a*) Sand. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The high-kurtosis heuristic $\beta_{high-kurt}$ was used……...85

4.4 (*a*) Bark. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used…….……85

4.5 (*a*) Pigskin. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used…….……86

4.6 (*a*) Herringbone Weave. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions……..……86

4.7 (*a*) Straw. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions…….87

4.8 (*a*) Grass. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used…….…87

4.9 (a)-(d) Examples of frequency responses of MICA filters corresponding to……92 The Gravel texture; (e) magnitude $|G|$ of the MICA interaction matrix for the Gravel texture. The larger the magnitude of $G_{i,j}$ the greater the statistical dependency between the corresponding MICA components.

4.10 High-level MICA algorithm……..92

4.11 Function φ: Linear in the unit interval and quadratic outside……92

5.1 Depiction of (unit diameter) BD NANS window with semi-circular upper and lower halves with rotation and offset from an ideal straight-line boundary non-stationarity……116

5.2 NANS processing of a multi-texture image. (*a*) multi-texture image; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1…....117

5.3 NANS processing of a multi-texture image. (*a*) multi-texture image; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1…....118

5.4 NANS processing of a fingerprint image. (*a*) fingerprint; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1…….……..…..119

5.5 NANS processing of a fingerprint image. (*a*) fingerprint; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1…….....120

5.6 NANS processing of a natural image. (*a*) van Hateren image #1122 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown…….……121

5.7 NANS processing of a natural image. (*a*) van Hateren image #8 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown…….……122

5.8 NANS processing of a natural image. (*a*) van Hateren image #93 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown…….123

6.1 Non-stationary analysis of a (*a*), (*b*) fingerprint image (*c*), (*d*) a natural image..151

6.2 Comparison of texture-contrast with human fixations on van Hateren image #245. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……..…...152

6.3 Comparison of texture-contrast with human fixations on van Hateren image#37. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations…...…153

6.4 Comparison of texture-contrast with human fixations on van Hateren image #122. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……....154

6.5 Comparison of texture-contrast with human fixations on van Hateren image#34.

(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……..155
6.6 Comparison of texture-contrast with human fixations on van Hateren image #161. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……...….156
6.7 Comparison of texture-contrast with human fixations on van Hateren image #232. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……..…..157
6.8 Comparison of texture-contrast with human fixations on van Hateren image #146. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……..……...158
6.9 Comparison of texture-contrast with human fixations on van Hateren image #54. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……159
6.10 Comparison of texture-contrast with human fixations on van Hateren image #353. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……160
6.11 Comparison of texture-contrast with human fixations on van Hateren image #190. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations……..……..161
6.12 Comparison of different probability maps corresponding to Image #232
(a) Human fixation induced probability map (unit-width Gaussians)
(b) GAFFE fixation map (unit-width Gaussians)
(c) Texture-contrast map (unit-width Gaussians)
(d) Texture-contrast map (twice-width Gaussians)…….……162
6.13 Average KLD, 1-Foveal Width. Error bars indicate standard deviations…..**.....**168
6.14 Harmonic Mean KLD, 1-Foveal Width. Error bars indicate standard Deviations……169
6.15 Average KLD, 2-Foveal Width. Error bars indicate standard deviations…..**.....**170
6.16 Harmonic Mean KLD, 2-Foveal Width. Error bars indicate standard Deviations……171

Chapter 1

Introduction and Background

1.1 Motivation

The basic problems of computational vision fall under two broad categories. The first is the formation of computational models of visual perception and cognition, including the discovery and elucidation of the important sub-processes that constitute vision. The second is the formation of computational models of neural processing in visual systems, including the experimental identification of important components of the visual system, explicating their role, and how they fit into the entire architecture of the system—in short, characterizing the architectural properties of visual systems.

A fundamental sub-category within visual perception is that of understanding the system level design principles of the vision system, and their implications with respect to information processing and behavioral aspects of the visual system in response to natural stimuli. Achieving such a basic understanding about visual systems can potentially have far reaching consequences not only to the science of visual perception, but also to the design of artificial vision systems that efficiently extract information from their natural surroundings and thus to their ability to execute complicated behavioral tasks. This dissertation aims to broaden our understanding of this aspect of computational vision for the special case of the Human Visual System (HVS).

The particular behavioral aspect of the HVS that will occupy us in this dissertation, and that serves as a vehicle for achieving a deeper understanding into the statistical

structure of natural scenes and their implications to optimal information processing in the HVS, is that of (low-level) fixation selection in natural scenes. Fixation selection is a special case of general *visual search processes* in the HVS which we define as the task of extracting relevant visual information from natural scenes. The visual information contained in natural scenes is extracted and analyzed at different layers of processing in the HVS—from the low-level (i.e. 'front-end' processing of the HVS) corresponding, for example, to contrast and luminance information, to higher-levels (i.e. 'back-end' processing of the HVS) corresponding, for example, to information about object class structures and contextual information. Thus there is a one-to-one mapping between the layers of visual processing and the visual information decomposed at that level.

From the point of view of low-level visual processes, the notion of visual information is intimately tied to the characterization of the low-level statistics of natural stimuli which we call Natural Scene Statistics (NSS)—in such a setting, as we shall see in more detail in this dissertation, optimal visual search tasks are mathematically well-defined and thus can be precisely studied. As we move to higher levels of visual processing, where cognition plays a dominant role, visual search tasks assume different forms given the behavioral tasks at hand (that implicitly define an underlying notion of visual information)—examples of which include object recognition (in the case where we are searching for objects in the scene), and image segmentation (if the aim is to partition the image into semantically different regions).

This dissertation is concerned exclusively with low-level visual search tasks which, as mentioned above, are more amenable to precise characterizations compared to their higher-level counterparts like object recognition. Moreover since high-level visual

processes built upon lower-level ones, we believe that achieving a basic understanding of low-level visual processes can potentially be fruitful for understanding cognitive mechanisms in the HVS as well.

1.2 A Role for Information Theory in Low-level Fixation Selection

The operation of any complex system like the HVS requires powerful organizational principles that allow the system to successfully function in complex natural environments. In the 1950s, Attneave [1] and Barlow [2] proposed one such organizational principle which has come to be known as the *efficient coding hypothesis*. They posited that information theory can provide a link between environmental statistics and the properties of neural responses, in that the retina and other stages of the early visual system have evolved to develop efficient codes (i.e. in the least number of bits) for the information processed at the respective stages (given biological constraints at each stage such as the available number of neurons etc). Verifying the hypothesis entails not only the discovery of rich Natural Scene Statistics (NSS) models but also establishing precise quantitative relationships to neural coding procedures that purportedly optimize certain aspects of NSS. Doing so would precisely establish the nature of the duality between NSS and low-level HVS processes.

Given the scope and generality of this hypothesis, various modified and restricted versions of this efficient coding hypothesis have been proposed and verified by researchers [3]-[7]. More recently, work in the above two-fold research program of developing powerful theoretical models for NSS coupled with investigations into their implications for information processing in the HVS [8]-[13] have greatly advanced.

In this dissertation we, for the first time, attempt to extend the scope of Barlow's hypothesis to fixation selection in natural images. Our general hypothesis is that low-level visual fixations performed by the HVS in natural scenes are driven by the goal of maximally extracting visual information from the scene. We do not verify this hypothesis in its full generality but rather for the specific cases of *contrast* and *textural* information. To do so we develop novel statistical characterizations of NSS with respect to contrast and textural information and demonstrate how these can be effectively exploited in devising optimal fixation patterns which are shown to match very well with actual human fixations performed on those scenes.

The rest of this dissertation builds the tools and concepts necessary to formulate the beginnings of a unified information-theoretic approach to constructing a computation theory of low-level visual search processes in the HVS.

1.3 Related Research

1.3.1 Foveation

One of the important factors that of visual perception that motivates eye movement and visual fixations is that the HVS is a foveated visual system: the sampling density is highest at the point of fixation and gradually decreases from there [14]-[16]. At any fixation, the image acquired by the HVS contains less detailed information in the periphery. To acquire peripheral information at high resolution, the eye makes rapid ballistic movements – saccades. Conversely, foveation dramatically reduces the amount of information processed at each fixation.

Much of the prior work on foveation has focused on achieving computationally fast and accurate implementations of foveation [16]. This is of course very important in order to be able to efficiently simulate foveated vision systems which in turn can be used to test various fixations strategies devised to emulate the HVS. Given that foveation is an important aspect of visual perception, in Chapter 2 we analyze the foveated processing in a more general setting by introducing the concept of linear scale variant (LSV) signals, and demonstrate how linear post-processing of LSV signals can be approximated by their linear shift invariant equivalents under a broad set of conditions that are usually met in practice as detailed in Chapter 2. This result simplifies the analysis of LSV filtered signals since at each point in the image, an equivalent linear channel accounts for the resulting image structure.

1.3.2 Natural Scene Statistics Models

Central to our investigations into low-level visual search mechanisms is the characterization of important statistical properties of natural scenes—the exploitation of which can potentially yield efficient low-level search strategies for the visual system. The philosophical underpinning of this approach is rooted in the hypothesized duality between NSS and low-level visual processes as extrapolated from Barlow's original enunciation of the same.

There is however, at present, no way to ascertain *a priori* as to which image feature, when statistically characterized, will prove to be useful for understanding vision. This choice is largely based on empirical examination coupled with intuition. Furthermore, a full characterization of NSS will involve multi- (or perhaps infinite-) dimensional

probability densities involving complex relationships between the image features identified above. Owing to the *curse of dimensionality* [17-18], the numerical estimation of such a hypothetical probability density is likely to be very formidable. Therefore it is much more fruitful to make progress in a bottom-up fashion by identifying important image features that exhibit statistical regularity, followed by a possible synthesis of different such statistical studies to furnish a more complete understanding of the intricate structure of NSS. Furthermore, such an approach is likely to afford us much greater insight into the corresponding structure of low-level visual processes of the HVS. Accordingly there has been a lot of progress over the past decade into characterizing salient aspects of NSS which we now briefly review.

Luminance and contrast are the most important low-level stimuli encoded the HVS. Using Weber's definition of contrast viz. the pixel luminance normalized by average luminance about a local window, the so-called luminance encoding of the HVS is actually a form of contrast encoding wherein bigger window sizes used for computing the average 'local' luminance level [14, 19]. Nevertheless for the purposes of distinction between small and big local window sizes, in the following discussion we reserve the term contrast for the case where smaller local windows are to compute the mean luminance level about a given pixel. Given the physiological importance of these low-level stimuli, many researchers have characterized luminance and contrast statistics in natural images in various ways [20-25]. It has been discovered from these studies that the luminance distribution is approximately symmetric on a logarithmic axis and hence positively skewed on a linear scale [20-22] i.e. with respect to the mean luminance there is a predominance of dark pixels compared to bright ones. Contrast statistics, on the other

hand, have been measured using various definitions of contrast including RMS contrast and Michaelson contrast—these various distributions have been shown to furnish consistent results [22-25]. The variations of luminance and contrast tend to be large in natural images. It turns out that these variations in contrast and luminance features, tend to be roughly independent of each other in natural images—though various degrees of positive or negative correlations are observed conditional on different image textures over which the measurements are made [25]. The implication of all these findings to neural coding in image has been investigated in [20, 22, 24].

Color is another important feature of natural images whose statistical characteristics have been examined together with their implications for visual coding. For color statistics, the aim is to characterize the joint distribution of the reflection spectra corresponding to different wavelengths of light that are reflected from materials and from natural light sources. It has been discovered that the observed spectra can be characterized almost completely with respect to relatively few of the most significant PCA components of the joint histogram of the reflectance spectra. This is a partial explanation for why there are only three cone receptors for representing color in the HVS (although it turns out that three receptive cones are less than optimal). Furthermore, it turns out that the joint distribution of the logarithm of the cone responses in natural images is approximately Gaussian [26-29].

So far we have been considering one-dimensional image features such as contrast and color. A more comprehensive examination however requires a probabilistic characterization of the spatial structure of natural images. This can be examined in several ways. One illuminating approach is to characterize the Fourier power spectral

density of natural images [30-31] which has consistently been shown to obey a $1/f^n$ power law, where n is approximately unity. This means that the amplitude spectra of natural images are relatively scale invariant in that scaling all frequencies by a factor—induced for e.g. varying the viewing distance of the scene from the plane of the eye—does not have significant affect on the shape of the amplitude spectrum.

A more principled comprehensive approach to modeling the spatial statistics of natural images, however, is to characterize the joint probability distributions among filtered responses to images where the filterbank could either be chosen *a priori* or could be learned from the image statistics themselves. An important example of the former approach is the observation that the joint distribution between neighboring wavelet coefficients of natural images tend to follow a GSM (Gaussian Scale Mixture distribution) [32-33] wherein a simple divisive normalization procedure renders the resulting random variables as jointly Gaussian. Correspondingly it has been shown that weighted divisively normalized models of V1 neurons do provide accurate descriptions of observed physiological response profiles of neurons in early visual processing [34]. Further investigations have demonstrated that such weighted normalized filtered outputs tend to be statistically independent [35].

On the other hand, the filterbank can also be learned from the data itself. An important way this can be accomplished is to learn the optimal filters that account for the joint probability distribution corresponding to *MxM* samples of natural images. The sampling can either be performed randomly or densely. It has been demonstrated by vision researchers that performing Independent Component Analysis (ICA) on the image patches tend to yield basis function that are similar to cortical receptive fields. These

results are consistent with the observation mentioned above that the output responses of neurons in the early visual system tend to be statistically independent. Though ICA-based representations of image patch statistics tend to yield sparse representations of natural image data [10, 13, 36], PCA-based representations do not furnish such compact representations of spatial structure since the variance of the data is spread out over multiple principle axes—thus rendering PCA less useful in providing useful probabilistic descriptions of the spatial structure of natural images.

Though numerous useful formulations of NSS have been emerged in the literature as briefly summarized above, these do not yield direct insights into how to formulate optimal low-level visual search strategies. Consequently we develop new formulations of NSS in this dissertation—specifically for the cases of contrast and texture statistics—that enable us, as we describe in detail in the coming chapters, to directly formulate and deploy optimal low-level visual search algorithms for natural scenes. We now briefly survey important and relevant prior work in the area of visual search.

1.3.3 High-level visual search

Early studies on eye movements, conducted by the Russian psychologist Yarbus [37] in the 1950s and 60s, revealed that visual fixations are very much influenced by high-level factors such as the nature of the specific task being performed. These results point to the importance of understanding top-down mechanisms involved in visual fixations. Top-down approaches are popular in computer vision because the problem can be intuitively formulated in terms of high-level features of the object such as shape, spatial relationships between objects, and so on. Wixson [38] proposed an ‘indirect search’

strategy using spatial relationships between targets and its surroundings to first identify an intermediate object (associated with the target) that is easier to find and then search in that region for the target. Since knowledge about cognitive mechanisms employed by the HVS during visual search is limited, top-down approaches usually incorporate *ad hoc* assumptions regarding what features will be of interest during fixation mechanisms.

Beyond the confines of foveated visual systems, extensive work in high-level visual search has been conducted in the sub-field of object recognition, including the formulation novel similarity measures for matching objects of interest in images [39-42], algorithms for learning object classes [43-46], for exploiting local invariant features for object matching [47-49], techniques for building grammars from of local features learnt from object classes [50-52], and statistical methods of modeling contextual information in images [53-55]. Object recognition, as a field, at present largely consists of an eclectic collection of different techniques such as those cited above, and does not, as yet, possess coherent theoretical underpinnings due in part to the infancy of the field.

Another class of visual search problems is image segmentation which can be broadly defined as the process of partitioning the image into semantically different regions. Here again the ill-posed nature of the problem together with the amorphous notion of 'semantically similar' has made progress in the most general setting formidable. Nevertheless useful partial theories have emerged in the literature [56-60] which, together with advancements in object recognition, hold promise towards making significant inroads into deepening our understanding of the computational principles that underlie the cognitive notions of visual similarity.

We expect that important and sustained progress in the area of high-level visual search processes, such as those described above, will eventually enable significant breakthroughs towards strengthening the theoretical foundations of the computational aspects of the various cognitive mechanisms underlying vision.

1.3.4 Low-level visual search

Although the importance of high-level factors in visual search processes conducted by the HVS is undisputed, there is also ample evidence to demonstrate that a significant proportion of the fixations performed by the HVS is driven by low-level features. The sheer volume of human fixations performed—about 15,000 fixations/hour—makes it implausible that the HVS uses computationally intensive semantic scene information to make a majority of the fixations.

One of the most emphatic illustrations of the limitations imposed by low-level vision on performance in visual search was demonstrated in [61], wherein the role of low level features (also called *visual cues*) in visual search was assessed by measuring variations in discrimination performance. The influence of high level factors in search was minimized by using constrained experimental conditions (for e.g., two-alternate forced choice experiments with spatially and temporally localized targets). They show that search results as measured by accuracy and speed of performance are indeed influenced by low level factors such as loss of spatial frequency in the retina and contrast masking.

Bottom-up approaches to fixation selection assume that eye movements are probabilistically driven by low-level image structures. Proponents of this paradigm [62-64] propose computational models for human gaze prediction based on image processing

algorithms that accentuate image features that are deemed relevant. A few reported studies on automatic visual search have examined fixation selection based on features such as contrast, edges, object similarity [65] or combinations of randomized saliency and proximity factors [66]. In an interesting study, Privitera & Stark [62] used a suite of algorithms such as detecting the presence of symmetry, center surround regions in images that resemble receptive field profiles, wavelets, contrast, and edges- per-unit-area to predict points of interest in an image. They compared these predictions with human eye fixations. The comparison of the predictions and human eye movements was accomplished by analyzing their spatial/structural binding (location similarity) and temporal/sequential binding (order of fixations). They report that around 50% of their computed fixations matched those of human observers. A recent and more comprehensive study of low-level fixations was conducted by Rajashekar *et. al.* [67] wherein point-of-gaze statistical analysis of visual fixations was coupled with foveated analysis of fixation points. Extending previous work [68] done in a non-foveated fixation framework, the authors in [67] demonstrate that points corresponding to human fixations exhibit higher values of contrast, bandpass contrast, luminance and bandpass luminance on average as compared to random fixations performed on natural scenes. Furthermore, they proposed a simple fixation selection strategy (named GAFFE) that linearly combines saliency maps corresponding to all these features. The resulting GAFFE-based fixations [67] outperform random-based fixation strategies in natural images with respect to the correlation measure.

An underlying premise of the various bottom-up investigations of visual search processes is that high-level cognitive operations performed by the HVS are likely to be

prefaced by fundamental low-level information gathering operations inasmuch as accurate high-level statistical inferences can only be formed if the basic statistical knowledge about image structure has been gathered. This therefore points to the possibility that initial low-level fixation pattern might be subsequently replaced by fixation patterns induced by high-level cognitive mechanisms. Of course, contextual knowledge about the visual scene being presented to the HVS also plays a fundamental role in determining the exact nature of the fixation mechanisms involved. It is also possible, however, that the HVS performs complicated joint optimization procedures for determining optimal cue combinations which, in conjunction with the contextual information involved, in turn determine resulting fixation patterns. It is the complicated nature of these possible interactions that has made it difficult to separate and distill the various causative factors involved in the visual search processes.

We expect that important and sustained investigations into low-level visual processes will shed light on the various complex cue combination mechanisms involved in vision.

In spite of the important and impressive progress made above in the area of visual search, none of the previous works (either top-down or bottom-up approaches) tackle this problem from an information theoretic point of view. Our goal is to construct a low-level computational theory of low-level visual search processes based on an information theoretic approach to modeling visual fixations. This dissertation makes the beginning steps in this direction wherein we demonstrate how elegant solutions can emerge from this approach that not only afford deeper insight into the structure of natural scenes but

also yield considerable explanatory and predictive power into the observed properties of low-level fixation processes conducted by the HVS.

1.4 Dissertation Overview

The primary contributions of this dissertation are as follows:

1. Deriving theoretical bounds on the LSI approximation of LSV-filtered signals
2. A new characterization of the contrast statistics of natural images that has a direct implication into devising optimal contrast-based fixations
3. Discovery of a Multilinear ICA (MICA) decomposition for natural images from which a theoretical non-stationarity index is derived
4. Introducing the concept of non-stationarity indices as a new statistical measurement for gauging the extent of non-stationarity across natural images. A detailed study of practical non-stationarity measures motivated by the MICA-based theoretical non-stationarity index
    a. Optimal texture-based fixations that exploit the non-stationary structure of natural images
5. Construction of a low-level theory of visual search processes based on an information-theoretic approach to modeling low-level visual fixations: Evaluations of contrast-based, texture-based, and joint texture-contrast based fixation strategies to actual human fixations with comparisons to randomized and previous reported fixation strategies

The remainder of this dissertation is organized in the following manner:

Chapter 2: Approximating Filtered Scale-Variant Signals

We develop theorems that place limits on the point wise approximation of the responses of filters, both linear shift-invariant (LSI) and linear shift-variant (LSV), to input signals and images that are LSV in the following sense: they can be expressed as the outputs of systems with LSV impulse responses, where the shift-variance is with respect to the filter scale of a single prototype filter [69]. The approximations take the form of LSI approximations to the responses. We develop tight bounds on the approximation errors expressed in terms of filter durations and derivative (Sobolev) norms. Finally, we find application of the developed theory to defoveation of images, deblurring of shift-variant blurs, and shift-variant edge detection.

Chapter 3: Contrast Statistics of Natural Images: Fixation Selection by Minimization of Contrast Entropy

The human visual system combines a wide field of view with a high-resolution fovea and uses eye, head, and body movements to direct the fovea to potentially relevant locations in the visual scene. This strategy is sensible for a visual system with limited neural resources. However, for this strategy to be effective, the visual system needs sophisticated central mechanisms that efficiently exploit the varying spatial resolution of the retina. To gain insight into some of the design requirements of these central mechanisms, we have analyzed the effects of variable spatial resolution on local contrast in 300 calibrated natural images. Specifically, for each retinal eccentricity (which produces a certain effective level of blur), and for each value of local contrast observed at that eccentricity, we measured the probability distribution of the local contrast in the

unblurred image. These conditional probability distributions can be regarded as posterior probability distributions for the "true" unblurred contrast, given an observed contrast at a given eccentricity. We find that these conditional probability distributions are adequately described by a few simple formulas. To explore how these statistics might be exploited by central perceptual mechanisms, we consider the task of selecting successive fixation points, where the goal on each fixation is to maximize total contrast information gained about the image (i.e., minimize total contrast uncertainty). We derive an entropy minimization algorithm and find that it performs optimally at reducing total contrast uncertainty and that it also works well at reducing the mean squared error between the original image and the image reconstructed from the multiple fixations. Our results show that measurements of local contrast alone could efficiently drive the scan paths of the eye when the goal is to gain as much information about the spatial structure of a scene as possible [70-73].

Chapter 4: MICA—A Multilinear ICA Decomposition for Natural Image Modeling

We refine the classical ICA decomposition using a multilinear expansion of the probability density function of the source statistics [74-75]. In particular, we introduce a specific non-linear system that allows us to elegantly capture the statistical dependences between the responses of the Multilinear ICA (MICA) filters. The resulting multilinear probability density is analytically tractable and does not require Monte Carlo simulations to estimate the model parameters. We demonstrate the MICA model on natural image textures and envision that the new model will prove useful for analyzing non-stationarity natural images using models.

Chapter 5: Non-stationarity Measurement in Natural Images

We introduce the concept of image *non-stationarity indices* that quantify the degree of statistical non-stationarity across an image [76-77]. Our approach takes the view that since natural images are generally non-stationary, characterizing the non-stationary structure of images may yield useful insights into identifying regions of high information. The theoretical basis of our approach lies in a recent novel characterization of image patch statistics of natural scenes called Multilinear Independent Component Analysis (MICA). Using MICA, we develop a theoretical measure of non-stationary in natural scenes that we use to define a practical and computationally efficient non-stationarity index which we call the *Natural Image Non-stationarity* (*NANS)* Index. We employ the NANS Index to characterize the non-stationary structure of a variety of naturalistic images. We anticipate interesting applications of the NANS Index towards understanding image texture, visual content, and visual attention.

Chapter 6: Texture-Contrast Based Fixation Selection in Natural Images

We formulate and verify a Barlow-type hypothesis for fixation selection in natural images, where the fixation patterns are designed to maximally extract certain types of low-level visual information from the image [78]. After a brief overview of contrast statistics and optimal contrast-based fixation selection, we develop an optimum texture-based fixation selection algorithm based on a recent theory of non-stationarity measurement in natural images [76]. Thereafter we propose a simple coupling of the optimal texture-based and contrast-based fixation algorithms which exhibits robust

performance for fixation selection in natural images. The performance of the fixation algorithms are evaluated for natural images by comparison to randomized fixation strategies via actual human fixations performed on the images. The fixation patterns obtained outperform randomized, GAFFE-based [67], and Itti [64] fixation strategies in terms of matching human fixation patterns. These results also demonstrate the important role that contrast and textural information play in low-level visual processes in the HVS.

Chapter 7: Contributions and Discussions

We conclude the dissertation. We summarize the problems addressed and solved in this dissertation, together with a discussion of important problems that emerge as a result of this work—including the wider implications to the areas of computational vision and signal/image processing.

Chapter 2

Approximating Filtered Scale-Variant Signals

2.1 Introduction

Foveation, as explained in Chapter 1, is an important component involved in the visual perception of images processed by the HVS. In this Chapter we explore the properties of foveated signal processing in a more general setting—that of Linear Scale Variant (LSV) signals—in which we derive tight bounds for point-wise LSI approximations of linearly post-processed scale variant signals. These results simplify the analysis of LSV filtered signals since at each point in the image an equivalent linear channel accounts for the resulting image structure.

We consider the linear processing of n-dimensional signals of the form:

$$z_{\sigma(x)}(\boldsymbol{x}) = \frac{1}{[\sigma(\boldsymbol{x})]^n} \int_{\boldsymbol{R}^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x})]\, f(\boldsymbol{x}-\boldsymbol{a})\, d\boldsymbol{a} \tag{2.1}$$

where $\boldsymbol{R}^n$ are the n-dimensional reals, $\boldsymbol{x} = (x_1, \ldots, x_n)$, f: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}$ is a continuously-differentiable signal being filtered by the linear filter kernel g: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}$, and $\sigma(\boldsymbol{x})$: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}^+$ is a non-negative, shift-variant scale function. Later, we will also consider discrete-domain n-dimensional signals having form analogous to (2.1) with appropriate substitutions made in the definition of the signals and functions. Clearly, (2.1) may be regarded as a linear shift-variant filtering of the signal f by the kernel g, where the shift-variance is a result of allowing the scale function σ to vary with $\boldsymbol{x}$. We therefore refer to (2.1) as a *scale-variant filtering* or *scale-variant convolution* of the signal f. We make the

notation $g \otimes f$ to denote such a scale-variant convolution, to be distinguished from the usual shift-variant convolution notation $g * f$.

In the sequel it will also be understood that a *scale-variant signal* refers to a signal that can be written as (2.1) (or in the corresponding discrete form given in Section III). If $\sigma(\boldsymbol{x}) = \sigma$ = constant, then (2.1) takes the form of the familiar linear shift-invariant convolution

$$z_\sigma(\boldsymbol{x}) = \frac{1}{\sigma^n} \int_{\boldsymbol{R}^n} g(\boldsymbol{a}/\sigma)\, f(\boldsymbol{x}-\boldsymbol{a})\, d\boldsymbol{a} \tag{2.2}$$

Scale-variant signals of the form (2.1) appear in numerous applications, such as image and video foveation (where images are intentionally non-uniformly blurred) as a method of perceptual compression [66, 79-81]; and in modeling undesirable degradations in signals that are blurred non-uniformly, e.g. by coma [82-83]. Prior work on scale-variant signal processing has consisted mainly in applications to modeling biological vision systems [84-86] including fast implementation of foveation filtering [16, 87]. Scale-variant filtering has also been touched upon in the context of steerable filters [88-89], and log-polar representation of images [90]. Equations (2.1) and (2.2) also bear strong resemblance to the continuous wavelet transform [91-92] if interpreted as a function of scale.

Here we are concerned with linear filtering of signals modeled as the responses of scale-variant linear systems, *viz.,* can be written in the form (2.1). We will develop theorems that place limits on the approximation of the responses of filters, both linear shift-invariant (LSI) and linear shift-variant (LSV), to input signals and images that are LSV in the sense expressed by (2.1): that they can be expressed as the outputs of systems

with LSV impulse responses, where the shift-variance is with respect to the filter scale of a single prototype filter. The approximations take the form of LSI approximations to the responses. We develop theorems that express tight bounds on the approximation errors, expressed in terms of filter durations and derivative (Sobolev) norms.

This chapter is organized as follows. Section 2.3 develops the basic model of filtering scale-variant signals of the form (2.1) in continuous coordinates. Approximations are given for the outputs of linear systems, both LSI and LSV, to inputs modeled as (2.1). Tight bounds are developed for the approximation errors. Interestingly, some of these bounds depend on the dimensionality, *n,* of the signal. In Section 2.3, analogous approximations and bounds are discovered for signals defined on discrete coordinates. Section 2.4 finds application of the developed theory to several problems of interest, including defoveation of images, deblurring of shift-variant blurs, and shift-variant edge detection. The chapter concludes in Section 2.5.

2.2 Quasi-Invariant Filtering of Scale-Variant Signals

In this section we develop the continuous-domain notation and theory of approximating linearly filtered scale-variant signals. In the first part, we state the relevant theorems. Proofs are assigned to the Appendix. In the second part, we give analytic examples of the utility of the theoretical results. Numerical results using real signals are given in a later section.

A. Approximations and Theorems

We consider linear filtering of signals modeled as the responses of scale-variant linear systems, *viz.*, can be written in the form (2.1). Specifically we study functions of the form

$$q_{\sigma(x)}(\boldsymbol{x}) = \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \cdot \frac{1}{[\sigma(\boldsymbol{x}-\boldsymbol{b})]^n} \int_{\boldsymbol{R}^n} g\left[\boldsymbol{a}/\sigma(\boldsymbol{x}-\boldsymbol{b})\right] f(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{a})\, d\boldsymbol{a}\, d\boldsymbol{b} \tag{2.3}$$

which is the linear filtering (either LSI or LSV) of the scale-variant signal (2.1) by the kernel *h*.

We shall also be interested in the filtered function

$$\hat{q}_{\sigma(x_0)}(\boldsymbol{x}) = h(\boldsymbol{x}) * z_{\sigma(x_0)}(\boldsymbol{x}) = h(\boldsymbol{x}) * z_{\sigma}(\boldsymbol{x}) \,\big|_{\sigma = \sigma(x_0)}$$

$$= \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \cdot \frac{1}{[\sigma(\boldsymbol{x}_0)]^n} \int_{\boldsymbol{R}^n} g\left[\boldsymbol{a}/\sigma(\boldsymbol{x}_0)\right] f(\boldsymbol{x}-\boldsymbol{b}-\boldsymbol{a})\, d\boldsymbol{a}\, d\boldsymbol{b} \tag{2.4}$$

This is the LSI convolution of the filter *h* with the function $z_{\sigma(x_0)}(\boldsymbol{x})$, but with the scale function held constant: $\sigma(\boldsymbol{x}) = \sigma(\boldsymbol{x}_0) = $ constant; hence $z_{\sigma(x_0)}(\boldsymbol{x})$ is space-invariant and (2.4) is a true double convolution. As indicated by (2.2), this is also the LSI convolution of *h* with (2.2), where $\sigma = \sigma(\boldsymbol{x}_0)$.

The coordinate $\boldsymbol{x}_0$ is the point at which we make approximation to the filtered scale-variant signal (2.3). In fact, the approximation is (2.4). Thus define the *quasi-invariant approximation* of $q_{\sigma(x)}(\boldsymbol{x})$ at the point $\boldsymbol{x}_0$:

$$\hat{q}_{\sigma(x_0)}(\boldsymbol{x}) \approx q_{\sigma(x)}(\boldsymbol{x}) \,\big|_{x = x_0}. \tag{2.5}$$

If the filter h and the scale variant signal $z_{\sigma(x_0)}(\boldsymbol{x})$ are such that the approximation (2.5) is close (in some sense), then (2.3) will be referred to as a *quasi-invariant filtering* of the scale-variant signal.

The following Theorem places a bound on the magnitude of the error

$$\varepsilon(\boldsymbol{x}_0) = q_{\sigma(x)}(\boldsymbol{x}) \mid_{x = x_0} - \hat{q}_{\sigma(x_0)}(\boldsymbol{x}) \mid_{x = x_0}. \tag{2.6}$$

of approximation (2.5). It is expressed in terms of the filter durations and certain derivative or smoothness norms of the signal being filtered and of the scaling function. We define these first.

The n-dimensional vector $\boldsymbol{\Delta g} = (\Delta g_1, \ldots, \Delta g_n)^T$ has elements

$$\Delta g_i = \int_{\boldsymbol{R}^n} v_i^2 g^2(\boldsymbol{v})\, d\boldsymbol{v} \;\; ; i = 1, \ldots, n \tag{2.7}$$

with identical definition for the Δh_i ; $i = 1, \ldots, n$ of $\boldsymbol{\Delta h}$. The elements Δg_i and Δh_i are the energy variances (durations) of $g(\boldsymbol{x})$ and $h(\boldsymbol{x})$ along the direction of the axis x_i; $i = 1, \ldots, n$.

The vectors $\boldsymbol{\delta f} = (\delta f_1, \ldots, \delta f_n)^T$ and $\boldsymbol{\partial\sigma} = (\partial\sigma_1, \ldots, \partial\sigma_n)^T$ have elements

$$\delta f_i = \int_{\boldsymbol{R}^n} \left| \nabla f_i(\boldsymbol{r}) \right|^2 d\boldsymbol{r} \;\; ; i = 1, \ldots, n \tag{2.8}$$

$$\partial\sigma_i = \int_{\boldsymbol{R}^n} \frac{|\nabla\sigma_i(\boldsymbol{b})|^2}{[\sigma(\boldsymbol{b})]^n} d\boldsymbol{b} \;\; ; i = 1, \ldots, n, \tag{2.9}$$

where $\nabla f_i(\boldsymbol{x}) = \partial f(\boldsymbol{x})/\partial x_i$ is the ith element of the gradient vector of $f(\boldsymbol{x})$. The elements δf_i and $\partial\sigma_i$, $i = 1, \ldots, n$ are derivative functionals, or Sobolev norms, which are measures of the smoothness of the functions $f(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})$ along the direction of the axis

x_i; $i = 1, \ldots, n$. The integrands of the functionals $\partial\sigma_i$ are weighted by the reciprocal of the scaling function, and so express a greater sensitivity when $\sigma(\boldsymbol{x})$ is small. Finally, given vectors length-n vectors $\Delta\boldsymbol{g}$ and $\delta\boldsymbol{f}$, we denote the vector inner product by $\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f}$.

Theorem 2.1 – When $n \neq 2$, the absolute error $|\varepsilon(\boldsymbol{x}_0)|$ is bounded from above as:

$$|\varepsilon(\boldsymbol{x}_0)| \leq \left| \frac{2}{2-n} \right| \sqrt{\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f}} \cdot \sqrt{\Delta\boldsymbol{h} \bullet \partial\sigma} \,. \quad \clubsuit \tag{2.10}$$

A number of comments regarding this result are in order. First, the bound is tight. As the filters $g(\boldsymbol{x})$, $h(\boldsymbol{x})$ are taken arbitrarily narrow, the RHS of (2.10) vanishes. Likewise, if the variation in the signal $f(\boldsymbol{x})$ is sufficiently small, then (2.10) becomes arbitrarily small: zero if the signal is constant. Likewise, if $\sigma(\boldsymbol{x})$ is made sufficiently smooth, then the bound vanishes.

Secondly, the weighted functional $\partial\sigma$ is of particular interest. At locations $\boldsymbol{x}_0$ where the scaling function $\sigma(\boldsymbol{x})$ becomes small, then the bound can become large unless the scale-variant filter or process is such that $\sigma(\boldsymbol{x})$ changes very slowly near $\boldsymbol{x}_0$. For example, in image foveation [66, 79-81], $\sigma(\boldsymbol{x})$ increases away from a presumed point of visual fixation (heavier blurring), but may be small near the point of fixation. Theorem 2.1 implies that at such points, slow changes in $\sigma(\boldsymbol{x})$ are required in order that the approximation (2.5) might hold accurately. This might be desirable to be able to, e.g., construct an algorithm for de-foveation (as we shall see).

Thirdly, the dependence on the dimensionality n of the involved signals is interesting. For signals of high dimensionality, it appears that the bound (2.10) becomes very narrow

– although the involved products may be larger in practice. For $n = 2$, the bound is useless! Hence the result seems of reduced interest for two-dimensional signals and images. However, a Corollary result will be given next that provides a useful bound even for $n = 2$. Moreover, the bound (2.10) can be applied to two-dimensional signals when the involved filters are separable, e.g., Gaussian.

First, a few more notations are required. Let

$$\delta f_{\max} = \max_i \delta f_{i,\max} \tag{2.11}$$

$$\delta f_{i,\max} = \sup_{\boldsymbol{R}^n} \left| \nabla f_i(\boldsymbol{r}) \right| \tag{2.12}$$

and define the alternate directional duration measures

$$Dg_i = \int_{\boldsymbol{R}^n} v_i^2 |g(\boldsymbol{v})| d\boldsymbol{v} \;\; ; i = 1, \ldots, n \tag{2.13}$$

which are the usual (non-normalized) function variances, and the (directionless) overall duration

$$Dg = \sum_{i=1}^{n} Dg_i \; . \tag{2.14}$$

Corollary 2.1 - The absolute error $|\varepsilon(\boldsymbol{x}_0)|$ is bounded from above as

$$|\varepsilon(\boldsymbol{x}_0)| \le n \sqrt{C_g \cdot C_h} \; \delta f_{\max} \; \delta\sigma_{\max} \; \sqrt{Dg \cdot Dh}$$

where $C_g = \int_{\boldsymbol{R}^n} |g(\boldsymbol{a})| d\boldsymbol{a}$, $C_h = \int_{\boldsymbol{R}^n} |h(\boldsymbol{a})| d\boldsymbol{a}$ ♣ (2.15)

This second result is also dimension-dependent, but with a different (linear) dependence on the dimension n and finite bound when $n = 2$. For large n, (2.15) may prove less useful than (2.10). The bound is again tight in all variables, becoming arbitrarily small as the filter durations are reduced, or as the filter or scale function are made sufficiently smooth.

Making comparisons between the bounds in Theorem 2.1 and Corollary 2.1 is difficult, since the durations and the smoothness measures all have distinct definitions, and each contains four terms that behave independently. Nevertheless, the two results substantiate one another, since both indicate that the quasi-invariant filtering approximation of scale-variant signals will tend to be accurate if the involved filters are of short duration, and if the filtered signal is smooth, and if the change in scale is not too rapid. These observations will be born out later in the numerical simulations (Section 2.4).

B. Illustrative Examples

We now examine a few interesting and illustrative examples. These were selected for their general significance and applicability, as opposed to the numerical simulations given later, which demonstrate specific examples of interest.

Signal Differentiation:

In numerous applications it is of interest to differentiate a signal, possibly following a linear (and perhaps scale-variant) filtering. For example, in image processing, directional derivatives, gradients or Laplacian operators highlight sustained intensity changes, or edges [93]. In many other applications, derivative operators highlight sudden changes or

signal transients, reveal trends, or when combined with nonlinear operations, demodulate AM-FM signals [94].

Suppose that we are given a scale-variant signal (2.1) with $n = 1$. In the context of what is to follow, this would usually be a signal that has been filtered with a linear low-pass (smoothing) function g, such as a Gaussian, with a varying scale parameter. Suppose then that the scale-variant signal (2.1) is passed through a k-fold differentiator, with impulse response

$$h(x) = \delta^{(k)}(x) = \frac{d^k}{dx^k}\,\delta(x). \tag{2.16}$$

In this case, the quasi-invariant approximation is

$$\theta_{\sigma(x_0)}(x) = \delta^{(k)}(x) * z_\sigma(x)\,\Big|_{\sigma = \sigma(x_0)}$$

$$= \frac{1}{[\sigma(x_0)]^n}\, g[a/\sigma(x_0)] * f^{(k)}(x-a), \tag{2.17}$$

the convolution of the scale-variant filter with the k^{th} derivative of f. If the variation in f and σ are finite, as measured by $\delta f_{\max}$ and $\delta\sigma_{\max}$, then (2.17) is *exact* when $k \neq 2$, since (2.15) is zero: $Dh = 0$. When $k = 2$, then $Dh = \infty$! Hence the bound is not applicable, although the approximation remains exact. In this example, (2.10) is also not applicable for any k, since the square of the generalized function (2.16) is not properly defined [95].

Inverse Filtering:

A basic, yet difficult operation in signal processing is the restoration of a signal that has been degraded by a linear distortion function, e.g., image blurred by defocusing or other undesirable smoothing function [96]. The problem is variously called restoration, deconvolution, or inverse filtering, depending on the details of the formulation. The

problem is complicated by frequency-domain zeros in the blur function, noise, and other uncertainties that leave the problem generally ill-posed. Yet even more difficult is the case where the blur function is shift-variant, viz., the degree or the nature of the blur changes from point to point or moment to moment [82-83]. This problem has been given only a small amount of attention, especially as compared to the case of shift-invariant linear distortion. However, it is of interest for many applications.

We consider the case of attempting to reverse a linear distortion that is scale-variant in the sense of (2.1). Our approach is to apply an "inverse filter" that is also scale-variant, referred to as *scale-variant inverse filter*. We first consider the noise-free case where the scale-variant linear distortion is the only degradation of the signal.

Model a signal distorted by n-dimensional scale-variant distortion function g using (2.1). Also assume, for simplicity, that the distortion is suitably well behaved, in the sense that g possesses no frequency-domain zeroes, although this is a practical impossibility. In the future, modifications of the example solution proposed here could be developed, e.g., pseudo-inverse solutions, etc. Denote the Fourier transform of g by

$$G(\boldsymbol{\Omega}) = \Im\{g(\boldsymbol{x})\} = \int_{\boldsymbol{R}^n} g(a)\exp(-j\boldsymbol{\Omega}^T\boldsymbol{a})d\boldsymbol{a} \tag{2.18}$$

At each $\boldsymbol{x}_0$, $f(\boldsymbol{x})$ is modified by taking the inner product of f with the scaled filter function

$$g_{\sigma(x_0)}(\boldsymbol{x}) = \frac{1}{[\sigma(\boldsymbol{x}_0)]^n} g\left[\boldsymbol{a}/\sigma(\boldsymbol{x}_0)\right], \tag{2.19}$$

which has Fourier transform $G_{\sigma(x_0)}(\boldsymbol{\Omega}) = G\left[\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{\Omega}\right]$. Then, define the *scale-variant inverse filter*

$$h_{\sigma(x_0)}(\boldsymbol{x}) = \Im^{-1}\left\{1/G\left[\sigma(\boldsymbol{x}_0)\boldsymbol{\Omega}\right]\right\}. \tag{2.20}$$

This idea is conceptually simple; at each coordinate, define the scale-variant inverse filter to be the inverse Fourier transform of the reciprocal of the Fourier transform of the filter kernel g scaled by the scaling function evaluated at the current point of interest, $\boldsymbol{x}_0$. The idea is that near $\boldsymbol{x}_0$, the signal has been modified sufficiently similarly to LSI filtering with $g_{\sigma(x_0)}(\boldsymbol{x})$ that the restoration will be accurate. We note that if $H_{\sigma(x_0)}(\boldsymbol{\Omega}) = [1/G_{\sigma(x_0)}(\boldsymbol{\Omega})]$ is not square integrable, then $H_{\sigma(x_0)}(\boldsymbol{\Omega})$ is a power-type signal, with an appropriate interpretation for its inverse Fourier transform expressed in terms of generalized functions.

We would be surprised if this idea for shift-variant inverse filtering has not been considered previously; it may even be quite old. However, we have been unable to find a single reference to such a method. In any case, Theorem 2.1 and Corollary 2.1 directly address the validity of the approach. The signal f should not vary too quickly; at points where it does, the bounds will be large and the approximation poor. Likewise, the rate of change of the scale of the LSV degradation (and hence of the restoration filters) should be small. Where it changes quickly, expect a poor approximation. The remaining question address the degradation filter durations, and the restoration filter durations. The questions are linked since one defines the other.

Theorem 2.1 addresses this question with some generality. The bound (2.10) is reduced if $\Delta\boldsymbol{g}_{\sigma(x_0)}$, $\Delta\boldsymbol{h}_{\sigma(x_0)}$ are both small; however, there are limits on how well this can be accomplished. Dropping the question of scale for simplicity, consider distortion $g \leftrightarrow$

G and inverse filter $h \leftrightarrow H$. By the Fourier transform frequency differentiation theorem and Parseval's formula:

$$\Delta g_i = \int_{\boldsymbol{R}^n} v_i^2 g^2(\boldsymbol{v})\, d\boldsymbol{v} = \frac{1}{2\pi} \int_{\boldsymbol{R}^n} \left| \frac{\partial}{\partial \Omega_i} G(\boldsymbol{\Omega}) \right|^2 d\boldsymbol{\Omega} \quad \text{for } i = 1, \ldots, n \tag{2.21}$$

$$\Delta h_i = \int_{\boldsymbol{R}^n} v_i^2 h^2(\boldsymbol{v})\, d\boldsymbol{v} = \frac{1}{2\pi} \int_{\boldsymbol{R}^n} \left| \frac{\partial}{\partial \Omega_i} H(\boldsymbol{\Omega}) \right|^2 d\boldsymbol{\Omega}$$

$$= \frac{1}{2\pi} \int_{\boldsymbol{R}^n} \left| \frac{1}{G(\boldsymbol{\Omega})} \right|^4 \left| \frac{\partial}{\partial \Omega_i} G(\boldsymbol{\Omega}) \right|^2 d\boldsymbol{\Omega} \quad \text{for } i = 1, \ldots, n \tag{2.22}$$

While short-duration linear degradation functions might often be encountered in practice, and so (2.21) might be small, the problem that arises is expressed well by (2.22): the duration of h is controlled by the reciprocal of G. Low-pass blur functions that completely or nearly eradicate high frequencies will have large durations, hence (2.10) will grow quite large. This is a new interpretation of the main limitation of inverse filters: excessive and unpredictable amplification of high signal frequencies, especially when noise is present. In this case, it limits the reversibility of scale-variant linear degradations and blurs. A consequence of this is that, in cases where scale-variant blurs are intentionally applied to signals, and are desired to be reversible, then the square of the Fourier transforms of the blur functions should not vanish faster than their derivatives.

If the scale-variant blur is accompanied by additive noise, then it is natural to define a scale-variant minimum mean-squared error (MMSE or Wiener) filter, by applying the appropriate MMSE filter at each point in the signal.

Scale-Variant Random Process:

Suppose that f is a wide-sense stationary (WSS) random process $f(\boldsymbol{x}) = \tilde{f}(\boldsymbol{x})$ with mean μ_f and autocorrelation function $R_f(\boldsymbol{\xi}) = \mathrm{E}\left[\, \tilde{f}(\boldsymbol{x})\ \tilde{f}(\boldsymbol{x}-\boldsymbol{\xi}) \,\right]$. The scale-variant filtering (2.1) delivers a random signal $\tilde{z}_{\sigma(\boldsymbol{x})}(\boldsymbol{x})$ that is no longer WSS.

The mean function of the filtered process is

$$\mu_z(\boldsymbol{x}) = E\left[\frac{1}{[\sigma(\boldsymbol{x})]^n} \int_{\boldsymbol{R}^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x})]\, \tilde{f}(\boldsymbol{x}-\boldsymbol{a})\, d\boldsymbol{a} \right] \tag{2.23}$$

$$= \mu_f\, \mu_g \tag{2.24}$$

where,

$$\mu_g = \frac{1}{[\sigma(\boldsymbol{x})]^n} \int_{\boldsymbol{R}^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x})]\, d\boldsymbol{a} = \int_{\boldsymbol{R}^n} g(\boldsymbol{a})\, d\boldsymbol{a} = \text{constant.} \tag{2.25}$$

Since the filters have constant area over scale, then $\tilde{z}_{\sigma(\boldsymbol{x})}(\boldsymbol{x})$ has constant mean function.

The autocorrelation function of the output process $\tilde{z}_{\sigma(\boldsymbol{x})}(\boldsymbol{x})$ is

$$R_z(\boldsymbol{x}, \boldsymbol{\xi}) = E[\, \tilde{z}_{\sigma(\boldsymbol{x}-\boldsymbol{\xi}/2)}(\boldsymbol{x}) \tilde{z}_{\sigma(\boldsymbol{x}+\boldsymbol{\xi}/2)}(\boldsymbol{x}) \,]$$

$$= \frac{1}{[\sigma(\boldsymbol{x}-\boldsymbol{\xi}/2)\sigma(\boldsymbol{x}+\boldsymbol{\xi}/2)]^n} \int_{\boldsymbol{R}^n} \int_{\boldsymbol{R}^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x}-\boldsymbol{\xi}/2)]\, g[\boldsymbol{b}/\sigma(\boldsymbol{x}+\boldsymbol{\xi}/2)]\, E\left[\, \tilde{f}(\boldsymbol{x}-\boldsymbol{\xi}/2-\boldsymbol{a})\, \tilde{f}(\boldsymbol{x}+\boldsymbol{\xi}/2-\boldsymbol{b}) \right] d\boldsymbol{a}\, d\boldsymbol{b}$$

$$= \frac{1}{[\sigma(\boldsymbol{x}-\boldsymbol{\xi}/2)\sigma(\boldsymbol{x}+\boldsymbol{\xi}/2)]^n} \int_{\boldsymbol{R}^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x}-\boldsymbol{\xi}/2)] \int_{\boldsymbol{R}^n} g[\boldsymbol{b}/\sigma(\boldsymbol{x}+\boldsymbol{\xi}/2)]\, R_f(\boldsymbol{\xi}-\boldsymbol{b}+\boldsymbol{a})\, d\boldsymbol{b}\, d\boldsymbol{a} \tag{2.26}$$

the inner integral of which is a scale-variant convolution of the form (2.1). It is of interest to learn whether a useful approximation to (2.26) can be developed. The outer integral is

not a scale-variant convolution; therefore we cannot apply Theorem 2.1 or its Corollary to develop an approximation to (2.26). However, in a moment we shall state and prove a Lemma that will serve this purpose.

In fact we propose the approximation

$$\hat{R}_z(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{1}{[\sigma(\boldsymbol{x})]^{2n}} \left\{ g[\frac{-\boldsymbol{\xi}}{\sigma(\boldsymbol{x})}] * g[\frac{\boldsymbol{\xi}}{\sigma(\boldsymbol{x})}] * R_f(\boldsymbol{\xi}) \right\} \tag{2.27}$$

to $R_z(\boldsymbol{x}, \boldsymbol{\xi})$ which is expressed in terms of shift-invariant convolutions. The autocorrelation approximation (2.27) is still a function of position $\boldsymbol{x}$; *viz.,* from point-to-point in the signal, the (approximated) correlation structure changes. From a computational perspective, the correlation approximation must be computed via a convolution at every point, but it has the advantage that it need not be computed as a separate operation for every $\boldsymbol{\xi}$ as well, unlike the true expression (2.26).

The validity of the approximation (2.27) is addressed by the following Lemma. We denote

$$\rho(\boldsymbol{x}, \boldsymbol{\xi}) = R_z(\boldsymbol{x}, \boldsymbol{\xi}) - \hat{R}_z(\boldsymbol{x}, \boldsymbol{\xi}) \tag{2.28}$$

$$C_g = \int_{\boldsymbol{R}^n} |g(\boldsymbol{a})| \, d\boldsymbol{a} \, . \tag{2.29}$$

Lemma 2.1 - The absolute error $|\rho(\boldsymbol{x}, \boldsymbol{\xi})|$ is bounded from above as

$$|\rho(\boldsymbol{x}, \boldsymbol{\xi})| \leq n \cdot C_g \cdot \delta R_{f,\max} \cdot \delta\sigma_{\max} \cdot Dg \cdot \sum_{j=1}^{n} |\xi_j| \, . \quad \clubsuit \tag{2.30}$$

This result suggests that the formula (2.27) is most useful for small correlation distances. Indeed, when $\boldsymbol{\xi} = \boldsymbol{0}$, the approximation is exact. Thus, the approximation captures the

second-order point statistics (variances) exactly. The approximation bound is also tight: the error becomes arbitrarily small when the correlation function R_f is sufficiently smooth, when the scaling function g changes slowly enough, and when the filter g is adequately narrow.

As an example of these concepts, we explore the idea of *scale-variant zero-crossing rates.* If $\tilde{f}(\boldsymbol{x})$ is Gaussian, then the output process is Gaussian as well. In the case of a one-dimensional signal, so that $\tilde{f}(\boldsymbol{x}) = \tilde{f}(x)$ and $R_f(\boldsymbol{\xi}) = R_f(\xi)$, the input process has a zero-crossing rate expressed by Rice's famous formula [97]:

$$\lambda_0 = \frac{1}{\pi} \sqrt{\frac{R_f''(0)}{R_f(0)}} \quad . \tag{2.31}$$

Here we postulate an expression for the shift-variant zero-crossing rate for the case of a one-dimensional scale-variant process $\tilde{z}_{\sigma(x)}(x)$ with approximate autocorrelation function (2.27). The approximate zero-crossing rate at each x is (naturally enough):

$$\lambda_0(x) \approx \frac{1}{\pi} \sqrt{\frac{\hat{R}_z''(x,0)}{\hat{R}_z(x,0)}} \quad , \tag{2.32}$$

where

$$\hat{R}_z''(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{1}{[\sigma(\boldsymbol{x})]^{2n}} \left\{ g'[\frac{-\boldsymbol{\xi}}{\sigma(\boldsymbol{x})}] * g'[\frac{\boldsymbol{\xi}}{\sigma(\boldsymbol{x})}] * R_f(\boldsymbol{\xi}) \right\} \tag{2.33}$$

We have found it difficult to develop an error analysis of the approximation (32), so it remains as a postulate. However, in the simulations, we explore the utility of the approximation for a practical application: zero-crossing based edge detection. While the approximation (2.32) is 1-D, can also be used to approximate zero-crossing rates along appropriate paths (such as image scan lines) in higher dimensional signals.

2.3 Discrete Formulation

We now develop results for the case of scale-variant discrete-domain signals filtered by linear filters (LSV or LSI). Consider n-dimensional discrete-domain signals of the form:

$$z_{k(\boldsymbol{m})}(\boldsymbol{m}) = \sum_{p \in \boldsymbol{Z}^n} g\left[\boldsymbol{p}/k(\boldsymbol{m})\right] f(\boldsymbol{m}-\boldsymbol{p}) = g \otimes f \tag{2.34}$$

where $\boldsymbol{Z}^n$ are the n-dimensional integers, $\boldsymbol{m} = (m_1, \ldots, m_n)$, f: $\boldsymbol{Z}^n \rightarrow \boldsymbol{R}$ is a discrete-domain signal filtered by g: $\boldsymbol{Z}^n \rightarrow \boldsymbol{R}$, and k: $\boldsymbol{Z}^n \rightarrow \boldsymbol{R}^+$ is a non-negative, shift-variant integer-valued scaling function. Whenever $\boldsymbol{p}/k(\boldsymbol{m}) \notin \boldsymbol{Z}^n$, then we take $g(\boldsymbol{p}/k(\boldsymbol{m})) = 0$. We also refer to (2.34) as a *scale-variant filtering* or *scale-variant convolution* of f. When $k(\boldsymbol{x}) = k$ = constant, then (2.34) becomes

$$z_k(\boldsymbol{m}) = \sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} g(\boldsymbol{p}/k)\, f(\boldsymbol{m}-\boldsymbol{p}) \ . \tag{2.35}$$

We are concerned with filtering signals of the form (2.34). Thus we study functions of the form

$$q_{k(\boldsymbol{m})}(\boldsymbol{m}) = \sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} h(\boldsymbol{p}) \sum_{\boldsymbol{r} \in \boldsymbol{Z}^n} g\left[\boldsymbol{r}/k(\boldsymbol{m}-\boldsymbol{p})\right] f(\boldsymbol{m}-\boldsymbol{p}-\boldsymbol{r}) \tag{2.36}$$

which is the linear (LSI or LSV) filtering of (2.34) by h. We further define

$$\hat{q}_{k(\boldsymbol{m}_0)}(\boldsymbol{m}) = h(\boldsymbol{m}) * z_{k(\boldsymbol{m}_0)}(\boldsymbol{m}) = h(\boldsymbol{m}) * z_k(\boldsymbol{m}) \Big|_{k = k(\boldsymbol{m}_0)}$$

$$= \sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} h(\boldsymbol{p}) \sum_{\boldsymbol{r} \in \boldsymbol{Z}^n} g\left[\boldsymbol{r}/k(\boldsymbol{m}_0)\right] f(\boldsymbol{m}-\boldsymbol{p}-\boldsymbol{r}) \ . \tag{2.37}$$

which has the same explanation as (2.4): it is the LSI convolution of h with $z_{k(\boldsymbol{m}_0)}(\boldsymbol{m})$, but with the scale function held constant: $k(\boldsymbol{x}) = k(\boldsymbol{x}_0)$ = constant. The point $\boldsymbol{m}_0$ is where we make approximation to (2.36); the *quasi-invariant* approximation is (2.37):

$$\hat{q}_{k(\boldsymbol{m}_0)}(\boldsymbol{m}) \approx q_{k(\boldsymbol{m})}(\boldsymbol{m}) \Big|_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{2.38}$$

Again, if filter h and scale variant signal $z_{k(\boldsymbol{m}_0)}(\boldsymbol{m})$ are such that (2.38) is close, then (2.36) is a *quasi-invariant filtering* of the scale-variant signal.

Corollary 2.2, which follows, bounds the absolute value of the error

$$\varepsilon(\boldsymbol{m}_0) = q_{k(\boldsymbol{m})}(\boldsymbol{m}) \Big|_{\boldsymbol{m}=\boldsymbol{m}_0} - \hat{q}_{k(\boldsymbol{m}_0)}(\boldsymbol{m}) \Big|_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{2.39}$$

The bound is again expressed in terms of the durations of the involved filters and derivative norms of the signal and scaling function. The discrete directional durations are given

$$Dg_i = \sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} r_i^2 \left|g(\boldsymbol{r})\right|; \; i = 1, \ldots, n \tag{2.40}$$

and the overall duration Dg is still given by (2.14). The overall discrete smoothness functional is

$$(\nabla f)_{\max} = \max_{i\in\boldsymbol{Z},\boldsymbol{s}\in\boldsymbol{Z}^n} \left\{\nabla_i f(\boldsymbol{s})\right\} \tag{2.41}$$

$$\nabla_i f(\boldsymbol{s}) = f(\boldsymbol{s}) - f(\boldsymbol{s} - \Delta_i) \tag{2.42}$$

with the vector $\Delta_i = (0, \ldots, 1, \ldots, 0)^T$ taking nonzero value only in the ith position.

Corollary 2.2 - The absolute error $|\varepsilon(\boldsymbol{m}_0)|$ is bounded from above as

$$|\varepsilon(\boldsymbol{m}_0)| \le n\,(\nabla f)_{\max}\,(\nabla k)_{\max}\,\sqrt{Dg \cdot Dh} \quad \clubsuit \tag{2.43}$$

The bound (2.43) is tight in all terms. For filters *g* and *h* taken arbitrarily narrow, the bound vanishes; for signal *f* and scaling function *k* taken arbitrarily smooth, it also vanishes.

2.4 Simulation Results

In this section we show several examples of the *quasi-invariant approximation* in simulation. We find practical application to two problems, both suggesting avenues for using the ideas developed here while also serving to exemplify limitations found in schemes based upon such approximations.

Defoveation:

We begin by demonstrating an application of quasi-invariant approximation for *defoveation*. Foveation can be modeled as a scale-varying filtering system [66, 79-81], where the scale of the filter increases away from the point of fixation according to some scaling function.

Figure 2.1 shows a 512x512 foveated Lena image with 4 distinct annular regions of filter scales that increase away from the fixation point (presumed to be the image center). The prototype filter used was a unit variance Gaussian filter. Figure 2.2 shows the defoveated image. The defoveation is performed using the simple scheme described in Section II-B (Inverse Filtering). Figure 2.3 shows the quasi-invariant approximation. This foveated example was designed to contain sharp discontinuities in the scaling function. It is not representative of a foveation process reflective of the human eye or as appropriate

for human consumption. In this case, the defoveation scheme performs poorly, as might be expected from (2.43).

Figure 2.4 shows a more representative foveated image using the same prototype filter. Here the foveation is mediated by a gradual change in the filter scale away from fixation. Since the filter scale function varies smoothly, the Sobolev norm of the scale function in (2.43) is small, hence the quasi-invariant approximation is more accurate, as seen in Figs. 2.5 and 2.6.

Now consider the case where there is foveation blur accompanied by additive white Gaussian noise (AWGN). Figure 2.7 shows a graded foveated image corrupted by AWGN, while Fig. 2.8 depicts the defoveated image using the scale-variant inverse filters defined above; Fig. 2.9 shows the defoveated image using MMSE versions of scale-varying inverse filters; clearly, the scale-variant Wiener filtered image (Fig. 2.9) has much less noise amplification than the scale-variant inverse filtered image (Fig. 2.8).

Zero-Crossing Rate Approximation:

As a second type of example, the plausibility of the hypothesis in (2.32) is demonstrated. Figure 2.10 depicts plots of the theoretical and the actual zero-crossing rates obtained by applying scale-variant Laplacian-of-Gaussian (LoG) bandpass filters to Gaussian white noise 1-D signals. Here the ZC rate is plotted against the value of σ (expressed in units of sample rate) for the Gaussian filter component of the LoG. As can be seen, the average "theoretical" ZC rate as computed from (2.32), (2.33) is in close alignment with the actual ZC rates computed from the scale-variant filtered signals.

As an example of more specific application, note that the ZCs of LoG-filtered images are commonly used for scale-dependent edge detection in images [93]. Figure 2.11 depicts a scale-variant LoG-filtered image and also the associated ZC map that was computed from it. Although the graded scale-variant LoG was applied to the image tessellated on Cartesian coordinates, the ZC rates were measured by performing a coordinate transformation into polar coordinates centered at fixation (so that contours of constant radius map to columns). The ZC rate was computed along each row, then the ZC rates across the rows was averaged. To compute the theoretical ZC rates from (2.32), (2.33) the theoretical rate was computed for each σ for each row, then these were averaged across rows. Figure 2.12 shows the plots of theoretical vs. actual ZC rates. It may be noted that the theoretical ZC rate underestimated the actual ZC rates in the images; this is likely due to nonstationarities and non-gaussianity in the Lena image.

The implication of these results are that the quasi-invariant approximation may be extended, with care, for extended applications such as ZC rate approximation in scale-variant signals. Such signals can occur, e.g., in foveated edge detection systems.

2.6 Conclusions

The analysis of the structural responses of systems that depart from the usual assumptions of linearity and/or shift-invariance generally poses significant problems owing to the loss of the principles of superposition and/or frequency-domain equivalence. Analyzing such systems requires either the development of new tools for analysis, which is usually quite difficult, or the use of approximations that relate the systems to other, more easily-analyzed systems. We have taken the second approach here, but we believe that the

approximations used are simple enough and sufficiently understandable to find extensive applications. This is particularly likely owing to the increased recognition of the multi-scale (and often scale-variant) structure that is found in signals and images of recorded natural phenomena, such as speech signals, images, and videos.

Fig. 2.1 Foveated image
(with 4 annular regions)

Fig. 2.2 Defoveated image

Fig. 2.3 Quasi-invariant approximation
to the defoveated image

Fig. 2.4 Graded foveated image

Fig. 2.5 Defoveation of graded foveated image

Fig. 2.6 Quasi-invariant approximation of graded defoveated image

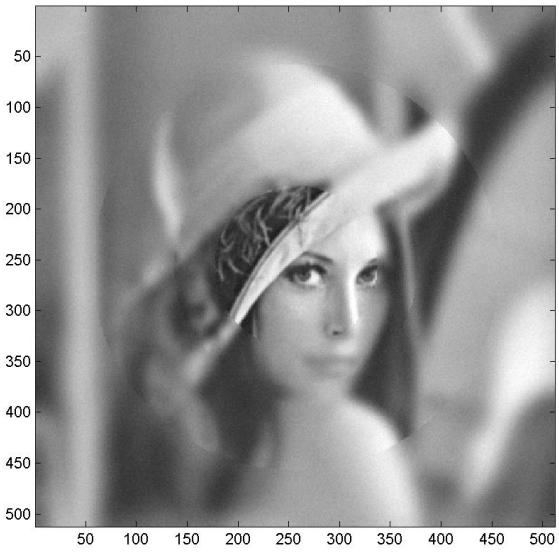


Fig. 2.7 Image foveated over four annular regions corrupted by AWGN (variance=10.0)

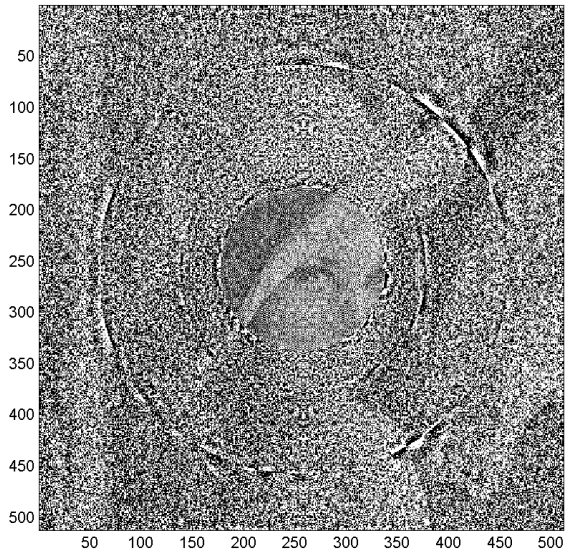


Fig. 2.8 Defoveated version of graded foveated noisy image in Fig. 2.7

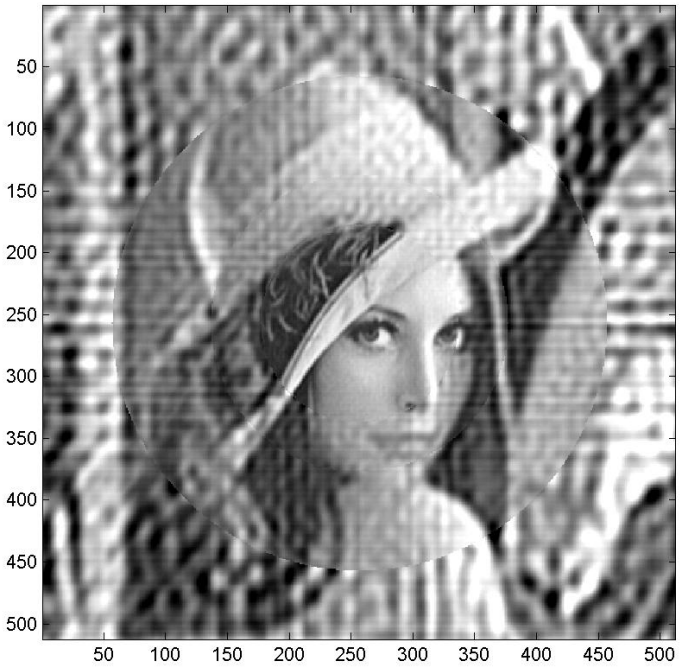


Fig. 2.9. MMSE (Wiener) defoveated version of graded, foveated noisy image in Fig. 2.7

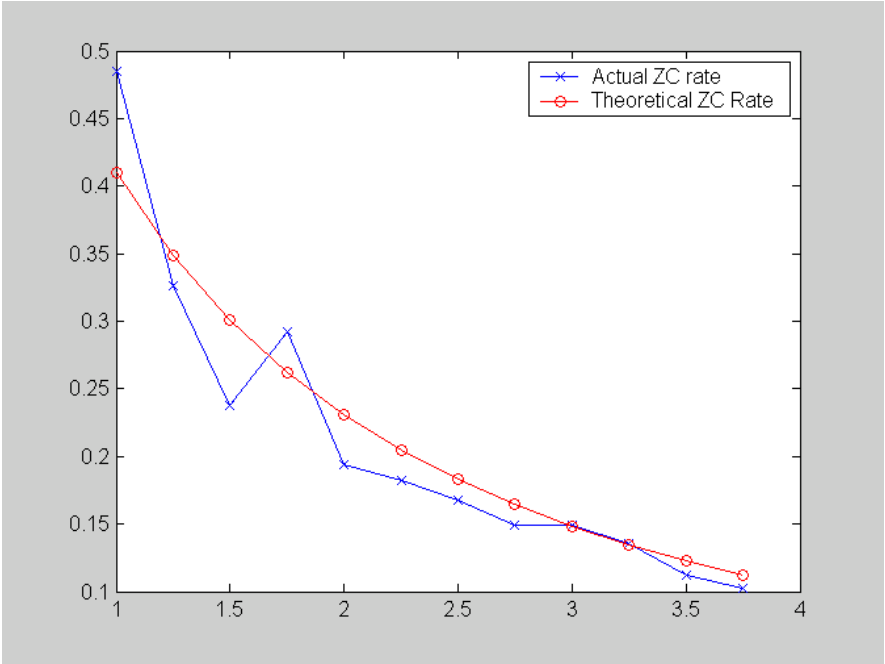


Fig.2.10. Plot of theoretical and actual zero-crossing rates averaged over 100 1-D Gaussian noise signals filtered by scale-variant linear Gaussian filters.

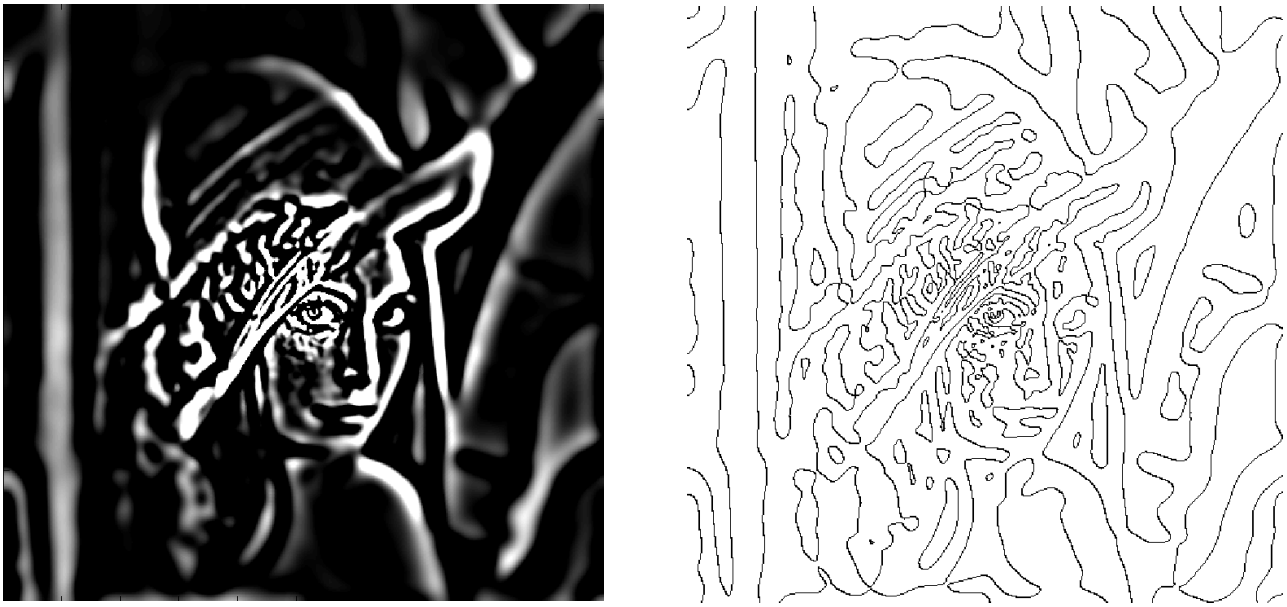


Fig. 2.11 Left: Scale-variant LoG-filtered Lena image. Right: ZCs computed from Left image

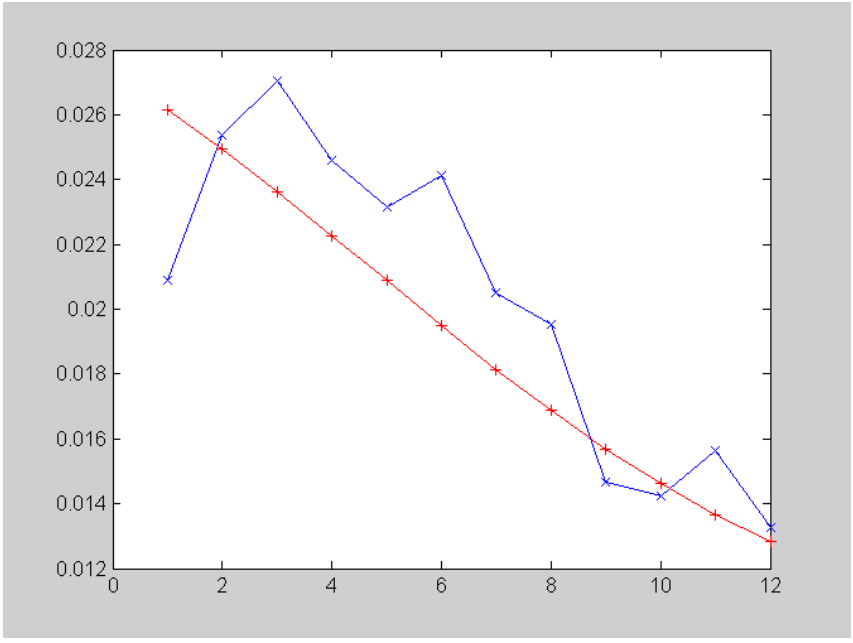


Fig. 2.12 Plot of theoretical and actual zero-crossing rates averaged over 100 radial directions on image filtered by scale-variant linear Gaussian filter.
(Legend: +:Theoretical ZC Rate, x: Actual ZC Rate)

Chapter 3

Contrast Statistics of Natural Images: Fixation Selection by Minimization of Contrast Entropy

3.1 Introduction

Humans, like many other animals, have a retina with variable spatial resolution. Resolution is highest in a central region, the fovea, and declines smoothly in all directions. High-speed eye movements, and slower head and body movements, are used to direct the fovea at potentially relevant locations in the retinal image of the visual scene. This strategy of combining a variable-resolution retina with eye, head, and body movements is sensible because it minimizes total neural resources while providing both a wide field view and high spatial resolution. However, for this strategy to be effective the visual system needs sophisticated central mechanisms that take into account and exploit the continuously varying spatial resolution of the retina.

There is evidence that visual systems are often matched to the statistical properties of the natural scenes to which they are exposed [10,13,20-21,30,98-103] (for reviews see Simoncelli and Olshausen [104] and Geisler and Deihl [105]). Therefore, to gain some insight into the design requirements of the central mechanisms of foveated visual systems, we analyzed the effects of variable spatial resolution on the statistics of local contrast in natural images. (Here, we define the local contrast as the standard deviation of the image intensities within some small region, divided by the mean intensity within that region, i.e., the local rms contrast.) Contrast is arguably the most fundamental local

image property encoded by the retina and transmitted to the brain, and hence its statistics have received considerable attention. A number of studies have been concerned with measuring the distributions of local contrast in natural images and comparing these with the shape of contrast response functions in the eye [20,21] lateral geniculate nucleus [24], and primary visual cortex [22,106]. Other studies have characterized the distributions of contrast in different environments [107] and at the center of gaze [68]. Like most other image properties, contrast is encoded with the greatest precision at the center of the fovea and with decreasing precision as the distance from the center of the fovea (the eccentricity) increases. Specifically, as eccentricity increases, the center sizes of ganglion cell receptive fields increase, blurring the retinal image and thereby effectively reducing local contrast and increasing contrast uncertainty. This fact motivated us to directly measure the effect of retinal blur on large numbers of natural images in order to determine the statistical relationship between effective contrast and the true unblurred contrast at different retinal eccentricities. Here, we show that to good approximation the mode ($\hat{c}$) of the posterior probability distribution of the unblurred contrast [i.e., the maximum *a posteriori* (MAP) estimate] is given by the simple formula:

$$\hat{c} = kc\varepsilon + c \tag{3.1}$$

where ε is the retinal eccentricity and k is a constant that depends on the patch size over which the local contrast (c) is computed. We also show that the average standard deviation (defined later) of the posterior probability distribution is given by:

$$(\bar{\sigma})^2 = (kc\varepsilon)^2 + \sigma_0^2 \tag{3.2}$$

where σ_0 is a small constant, and thus the contrast uncertainty (the differential entropy of the posterior probability distribution) is given by:

$$h = \frac{1}{2}\log_2\left(2\pi e(\bar{\sigma})^2\right) \tag{3.3}$$

These statistical properties of natural images will be derived and explained in Section 3.2.

As an example of how these statistical properties of natural images might be exploited by a foveated visual system, we have considered the task of selecting fixation locations, when the organism's goal is to encode images as well as possible with just a few fixations. Specifically, using Eqs. (3.1)–(3.3), we derive and evaluate a fixation selection strategy based on the principle of picking fixation locations that minimize the total uncertainty about the contrasts in the image (i.e., minimize the total contrast entropy). We decided to explore an algorithm that minimizes total contrast entropy because minimizing entropy is ideal under some circumstances and has proved useful in other applications [108-111]. We find that our algorithm works very well at reducing total contrast uncertainty and also works well at reducing the mean squared error (MSE) between the original image and the image reconstructed from the multiple fixations.

3.2 Methods and Results

This section describes the measurements of the contrast statistics and the algorithm for fixation selection based on those statistics.

3.2.1 Contrast Statistics

The effects of retinal blur on local contrast were measured using a set of calibrated natural images. The image set consisted of 300 rural images (i.e., minimum of man-made

objects or animals) obtained from a publicly available image database [13]. The images were selected to be as diverse as possible given the data set. The images were obtained with a Kodak DCS420 digital camera and were calibrated to result in approximately 12 bit values that are linear with respect to the luminance. The 1536 by 1024 images were cropped to the center 1024 by 1024 pixels. Van Hateren and van der Schaaf [13] report that each pixel corresponds to approximately 1 arc min, and thus the cropped images are approximately 17°x17°.

The contrast sensitivity functions of the human visual system, at different retinal eccentricities, have been measured for transient stimuli [112-114]. Measurements made under transient stimulus conditions are appropriate in the present context because fixation durations are brief (200–300 ms) under most natural viewing conditions. These contrast sensitivity functions are adequately described by the formula [16]:

$$C(f,\varepsilon) = C_0 \exp\left(-\alpha f \frac{\varepsilon_2 + \varepsilon}{\varepsilon_2}\right) \tag{3.4}$$

where α is a constant ($\alpha \cong 0.1$), ε_2 is the retinal eccentricity where spatial resolution falls to half of what it is in the center of the fovea ($\varepsilon_2 \cong 2.3^o$), and C_0 is a constant that controls the maximum contrast sensitivity. The contrast sensitivity functions described by Eq. (3.4) are consistent with the increase in center size of the retinal ganglion cells (midget ganglion cells) with eccentricity [113-114], and hence Eq. (3.4) can be used to estimate the reduction in effective contrast as function of eccentricity. Note that the blur produced by the retina (as reflected in ganglion cell center sizes) is a result of both optical and neural factors.

To simulate the blur produced by the retina at different eccentricities, we filtered each of the 300 natural images with radially symmetric transfer functions obtained by setting $C_0 = 1.0$ and $f = (f_x^2 + f_y^2)^{1/2}$ in Eq. (3.4). Specifically, for each image we padded it appropriately, took the Fourier transform, multiplied the result by Eq. (3.4), and then took the inverse Fourier transform. Blurred images were obtained for eccentricities (ε) of 0, 1, 2, 4, 8, and 16 deg. The filtered images at an eccentricity of 0 deg were taken to be the unblurred reference images. This was done because the optical transfer function of the camera is unknown (but presumably very good), and hence the raw image cannot be taken to be the effective retinal image in the center of the fovea. Because the unblurred image was taken to be the filtered image with $\varepsilon = 0$, the value of C_0 is irrelevant and hence could be set to 1.0, as we did.

In order to characterize the statistical relationship between effective contrasts at different eccentricities, we measured local contrasts in each image, for all six levels of blur. A large number of local contrasts were sampled randomly from each of the 300 natural images. The locations of the samples were different for each natural image but were the same for each level of blur. The local contrasts were measured in image patches formed by windowing with a circularly symmetric raised-cosine weighting function:

$$w_i = 0.5\left\{\cos\left[\frac{\pi}{p}\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}\right] + 1\right\} \tag{3.5}$$

where p is the patch radius, (x_i, y_i) is the location of the i^{th} pixel in the patch, and (x_c, y_c) is the location of the center of the patch. (Note that the half-height diameter of the window equals the patch radius.) The results reported here are for a patch diameter of

32 pixels (0.53 deg), but similar results are obtained with other patch sizes. The local contrast was defined by the formula:

$$c = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(L_i - L)^2}{(L + L_0)^2}} \tag{3.6}$$

where N is the number of pixels in the patch; L_i is the luminance of the i^{th} pixel; L is the local mean luminance:

$$L = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i L_i \tag{3.7}$$

and L_0 is a dark light parameter, chosen to be 7 td (1 cd/m2, assuming a 3 mm pupil), based on human photopic intensity discrimination data [115] (We note that L_0 had very little effect on the measured contrasts because the mean luminances of the images were generally much higher than 1 cd/m2.)

Figure 3.1 shows the estimated probability distributions of local contrast for each level of blur. The distributions have been truncated at a contrast of 0.005 because humans cannot detect contrasts below that value and because the measurements become contaminated by camera or pixel noise. Not surprisingly, as the level of blur (retinal eccentricity) increases, the distributions shift toward lower contrasts. The rise in the function at low contrasts appears to be due to the patches of sky in many of the natural images.

For many visual tasks (including the fixation selection task), one would like to estimate the unblurred contrast from the blurred contrast observed at the given retinal eccentricity. Thus, the statistics of most relevance are the conditional probability distributions for the unblurred contrast given the observed contrast (i.e., the posterior

probability distributions). We computed these distributions for a wide range of blurred contrasts, for eccentricities 1, 2, 4, 8, and 16 deg. Several representative distributions are shown in Fig. 3.2. Each row shows the conditional probability distributions for a different eccentricity, and each plot within a row shows the distribution for a particular value of blurred contrast observed at that eccentricity. There are several clear trends in the data: (1) As eccentricity increases, the peaks of the distributions shift to the right and (2) the widths of the distributions increase; and (3) as the observed blurred contrast increases, the peaks of the distributions shift to the right and (4) the widths of the distributions increase.

To quantify these trends, we fit the empirical distributions with descriptive functions. In general, the distributions are not Gaussian, but they are nicely fit by Gaussian distributions with different standard deviations above and below the mode (skewed Gaussian distributions). The solid curves show the fits to this sample of empirical distributions; the quality of these fits is representative of the whole set. The skewed Gaussian has three parameters: the mode, which we will label $\hat{c}$ because it is the MAP estimate of the unblurred contrast, and two standard deviations, σ_l and σ_h. Figure 3.3 plots the mode and the average standard deviation, $(\sigma_l + \sigma_h)/2$, for all eccentricities and observed levels of blurred contrast. Measurements outside these ranges were unreliable because the numbers of samples became too small. The solid lines in the figure are best fitting straight lines through the origin. Although the fits are not perfect, the straight lines summarize the data very well. In other words, to close approximation, both the mode and the standard deviation of all the posterior probability distributions increase in direct proportion to the observed blurred contrast.

What is also clear in Fig. 3.3 is that the slopes of the best-fitting lines (the proportionality constants) increase with retinal eccentricity. Figure 3.4 plots the estimated slopes for the modes and the average standard deviations. The straight line in Fig. 3.4A is the best-fitting line with an intercept of 1.0, and the straight line in Fig. 3.4B is the best-fitting line through the origin. Again, the fits are not perfect, but they do provide a very good summary of the data. Taken together, Figs. 3.2–3.4 show that the mode across all conditions is closely approximated by Eq. (3.1) and the average standard deviation across all conditions is closely approximated by Eq. (3.2), where $\bar{\sigma} = (\sigma_l + \sigma_h)/2$.

Differential entropy is a fundamental measure of the uncertainty associated with a probability distribution [116]. In Appendix B we show that the differential entropy of a skewed Gaussian distribution is equal to Eq. (3.3), and hence the differential entropy of the posterior probability distributions (the contrast uncertainty) for the range of eccentricities considered here is closely approximated by substituting Eq. (3.2) into Eq. (3.3). The constant σ_0^2 in Eq. (3.2) reflects the fact there must always be some intrinsic uncertainty about contrast, if for no other reason than photon and sensor or neural noise. Although the constant cannot be estimated from the contrast measurements, it is necessary for it to have a value greater than zero in the fixation selection algorithm; its specific value is not important as long as it is small (see Appendix A).

3.2.2 Fixation Selection

We have found a surprisingly simple statistical relationship, for natural images, between the contrast observed at a given retinal eccentricity and the posterior probability distribution of the unblurred true contrast at that location. This relationship, which is

described by Eqs. (3.1)–(3.3), could be exploited by a visual system to efficiently select fixation locations under certain circumstances. For example, if the goal in some situations is not to search for a particular target or set of targets but simply to gain as much information as possible about the image on each fixation, then a potentially effective strategy would be to pick successive fixations that maximally reduce the total contrast uncertainty about the image. This strategy might be particularly effective if there is a strong correlation between the uncertainty about local contrast and the total uncertainty about the local image structure. To begin exploring this possibility, we have developed an algorithm (a model observer) that selects fixations based on Eq. (3.1)–(3.3). Here, we describe the algorithm, then we describe the algorithm's fixation selections on some example images, and finally we compare the algorithm's absolute performance to appropriate ground-truth measurements.

3.2.2.1 Contrast Entropy Minimization Algorithm

We assume that the first fixation is at some arbitrary image location (e.g., at the center of the image). On making this first fixation the observer receives a foveated neural image, where spatial resolution is highest at the fixation point and falls off smoothly in all directions. From this first neural image the observer forms three maps that will be updated after each fixation. The first is an eccentricity map, which stores, for each image pixel, the smallest distance the pixel has been from the center of the fovea. The second is a contrast map, which stores the local rms contrast measured at each pixel, when the pixel was at its smallest distance from the center of fovea. (The contrast at a pixel is defined to be the contrast of the patch centered on that pixel.) The third is an uncertainty map, which

stores the contrast uncertainty (entropy) at each pixel [given by Eq. (3.3)], when the pixel was at its smallest distance from the center of fovea. These three maps cumulate all the relevant information obtained during the sequence of fixations. The sum of all the uncertainties in the uncertainty map is the total contrast uncertainty. The aim of the algorithm is to select the next fixation that will minimize this total contrast uncertainty. To do this, the algorithm considers every possible next fixation location. For each possible fixation location, the algorithm uses the current maps and its knowledge of the posterior probability distributions for contrast [Eqs. (3.1)–(3.3)] to estimate the reduction in total contrast uncertainty. It then picks the fixation location with the largest estimated reduction. A formal derivation of the contrast entropy minimization (CEM) algorithm is given in Appendix A.

3.2.2.2 Performance of the CEM Algorithm

The performance of the CEM algorithm was evaluated on 16 natural images selected to be representative of the van Hateren and van der Schaaf [13] data set. Thumbnails of these images are shown in Fig. 3.5. We simulated a foveated visual system that approximately matched the human visual system by using radially symmetric transfer functions corresponding to human contrast sensitivity functions [cf. Eq. (3.4)]:

$$F(f_x, f_y, \varepsilon) = \exp\left(-\alpha\sqrt{f_x^2 + f_y^2}\,\frac{\varepsilon_2 + \varepsilon}{\varepsilon_2}\right) \tag{3.8}$$

For each eccentricity the inverse Fourier transform of this transfer function specifies a linear filter kernel (a Laplacian function) that scales in size with eccentricity. To speed the calculations, we made use of the fact the resolution of the human visual system

declines smoothly as a function of eccentricity. By setting the left side of Eq. (3.8) to any constant resolution criterion, we see that resolution follows a smooth function of the form

$$r(\varepsilon) \propto \frac{\varepsilon_2}{\varepsilon_2 + \varepsilon}$$

The greatest eccentricity that needs to be considered for our 17° images is 12°, and hence the lowest relevant resolution is approximately 17% of the resolution in the fovea. Therefore, we partitioned the 17%–100% range into eight evenly spaced resolutions and then determined the eccentricity corresponding to each resolution. We then created eight transfer functions by substituting the eight eccentricities into Eq. (3.8). Before running the algorithm on a natural image, we used the eight transfer functions to obtain eight different resolution versions of the natural image. During the simulation, the foveated (neural) image at any given retinal eccentricity was obtained by linearly interpolating the two images whose resolutions bracketed the resolution at that eccentricity.

On each fixation during the simulation, the local contrasts in the neural image were measured using Eq. (3.6) for a patch diameter of 32 pixels. To speed the calculations, we sampled the local contrasts on a square lattice with a spacing of 16 pixels (the radius of the raised-cosine window). The overlap of the samples ensured that all image pixels contributed to the local contrast measurements (however, the algorithm performs similarly if there is no overlap between samples). The possible fixation locations and the three maps (contrast, eccentricity, and uncertainty) also corresponded to the same square lattice (i.e., 4096 possible fixation locations).

Figures 3.6A and 3.6C show the first nine fixations for two of the natural images. (Recall that the first fixation was always at the center of the image.) There are several trends evident in these fixation patterns. First, the fixations tend to land in or near

relatively high-contrast regions. Notice, for example, how there are no fixations into the sky region of the image in Fig. 3.6A and how the second fixation is near a bright flower in Fig. 3.6B. This occurs because contrast uncertainty is greater in regions where the effective contrast is higher [see Eqs. (3.2) and (3.3)]. Second, the saccade lengths tend to be relatively large and variable in size; the mean and standard deviation of the saccade length for the 16 test images are 8.9° and 2.5°, respectively. The large saccades occur because contrast uncertainty increases with eccentricity [see Eqs. (3.2) and (3.3)]. Third, there are few fixations near the edge of the image. This occurs because fixating near the image boundary tends to reduce the total number of image pixels that benefit from being seen at a smaller eccentricity. For example, a fixation on the boundary implies that half the fovea falls outside the image, which tends to reduce the number of image pixels that can benefit from foveal viewing.

Figures 3.6B and 3.6D show quantitatively how well the algorithm performs in reducing total contrast uncertainty. The solid circles show the total contrast entropy predicted by the algorithm before the fixation was made, where the total contrast entropy has been normalized by its value after the first fixation in the center of the image. The open circles show the actual total contrast entropy observed after the fixation selected by the algorithm is made. In other words, the predicted entropy is the entropy estimated before the next eye movement is made, and the observed entropy is the entropy observed or computed after the next eye movement is made. As can be seen, the predicted and observed entropies are very similar. The open triangles show the lowest possible total contrast entropy that could have been obtained on the fixation. It was determined by literally making every possible fixation and computing the observed entropy. The actual

observed entropy obtained by the algorithm is almost indistinguishable from optimal. The average results for all 16 images are shown in Fig. 3.7A. In general, the reduction in contrast entropy obtained by the CEM algorithm is essentially optimal. This is even more clearly illustrated by the solid circles in Fig. 3.7B, which plot the ratio of the optimal and observed entropies in Fig. 3.7A (the first fixation is excluded from the plot because the ratio is necessarily 1.0).

An obvious question is how well the CEM algorithm compares with alternatives. We consider two. The first algorithm tiles the image in a random order without replacement. Specifically, the image is divided into nine square regions (a 3x3 grid), and only fixations at the centers of these regions are allowed. During the scan, each square region is fixated only once, with the order of fixations being random. The average performance of this tiling algorithm is given by the open circles in Fig. 3.7B. It performs substantially worse than entropy minimization. The second alternative is purely random fixation (fixations are selected randomly from the 4096 possible locations). The performance of this algorithm is given by the open triangles in Fig. 3.7B. The random algorithm performs worse than the tiling algorithm. We conclude that the CEM algorithm does, in fact, optimally reduce the total contrast entropy on successive fixations for natural images and that it substantially outperforms some obvious alternatives.

We have demonstrated that the average contrast statistics of natural images can be used to sequentially select fixations that optimally reduce the total contrast uncertainty for individual images. Although this is a remarkable fact, contrast is just one local image property. Presumably, humans make fixations not just to reduce uncertainty about contrast but also to reduce uncertainty about many of the other image properties that

determine local image structure (e.g., orientation, phase, and spatial frequency). It is not possible to measure the statistics for all local image properties in natural images, and hence it is not practical to develop a rigorous algorithm that selects fixations to reduce total image uncertainty. On the other hand, it is possible that uncertainty in contrast is strongly correlated with uncertainty for other image properties. For example, Schwartz and Simoncelli [34] found that the variances of many local image properties are strongly correlated, even for orthogonal image properties. Therefore, it is possible that minimizing contrast uncertainty would do a good job of minimizing uncertainty about many local image properties.

To evaluate this possibility, we used the mean squared error (MSE), between the original (unblurred) image and the image reconstructed from the sequence of fixations, as a measure of the total image uncertainty. The reconstructed image was obtained using the eccentricity map (the map showing the smallest distance that each pixel has been from the center of the fovea). Specifically, each pixel in the reconstructed image was set to the image gray level that was observed at that pixel for the eccentricity given in the eccentricity map. Thus, in the reconstructed image, every pixel keeps the highest resolution that has occurred so far in the sequence of fixations. (We note that for image reconstruction the eccentricity map was computed for all the 1024x1024 pixels' locations in the image; also, the MSE between the original and reconstructed images was computed over all 1024x1024 pixels.)

For each fixation made by the CEM algorithm, we computed the relative MSE (the MSE after the fixation divided by the MSE after the first fixation). The solid circles in Fig. 3.7C show the relative MSE as a function of fixation number, averaged across the 16

test images. For ground-truth comparison, we determined, for each fixation made by the CEM algorithm, the fixation that would have minimized the MSE (this was done by making every possible next fixation and computing the resulting MSE). The open circles in Fig. 3.7C show the optimal values of the MSE that could have been obtained. The solid circles in Fig. 3.7D show the ratios of the optimal MSE to the observed MSE obtained with the CEM algorithm. The average ratio is 0.9 (i.e., the obtained MSE is about 10% higher than optimal). The open circles and triangles show that the tile and random algorithms perform considerably worse than the CEM algorithm; the average ratio for the tile algorithm is 0.72 and for the random algorithm is 0.59. Thus, it appears that the CEM algorithm does a respectable job of selecting fixations that minimize total image uncertainty.

3.3 Discussion

To gain insight into the design requirements of visual systems with foveated retinas, we measured the joint distribution of the local contrast in 300 natural images before and after blurring by amounts corresponding to different retinal eccentricities in the human visual system. The joint distribution at each retinal eccentricity is given by the marginal distribution of the blurred contrast (e.g., one of the distributions in Fig. 3.1) and by the distributions of the unblurred contrast conditional on the blurred contrast (e.g., one of the rows of distributions in Fig. 3.2). We find that the conditional distributions are described quite well by very simple formulas: The mode of the conditional distribution increases in proportion to the blurred contrast and the eccentricity [Eq. (3.1)], the average variance of the conditional distribution increases in proportion to the square of the blurred contrast

and the square of the eccentricity [Eq. (3.2)], and the differential entropy of the conditional distribution increases in proportion to the logarithm of the average variance [Eq. (3.3)].

The image statistics reported here are for one particular analysis patch size (a width of 32 pixels). We find that Eqs. (3.1)–(3.3) also summarize the conditional probability distributions for other patch sizes quite well. However, as patch size decreases, the estimated value of the proportionality constant k in Eqs. (3.1)–(3.3) increases.

To explore how these natural scene statistics might be exploited by central perceptual mechanisms, we considered the task of selecting successive fixation points to optimize the total contrast information gained about the image (i.e., minimize total contrast entropy). On the basis of the average scene statistics represented by Eqs. (3.1)–(3.3), we derived a novel fixation selection algorithm: the CEM algorithm. Remarkably, we found that the average scene statistics for natural images (represented in the CEM algorithm) are sufficient to achieve nearly optimal fixation sequences for individual natural images (see Figs. 3.6, 3.7A, and 3.7B). Presumably, this optimal performance is achieved because each fixation is based on a global pooling of local contrasts from the entire image. In other words, even though there is considerable uncertainty about how much the contrast entropy will be reduced at any particular image location, there is little uncertainty about how much the average contrast entropy from all locations will be reduced. We also examined how well the CEM algorithm performed at reducing the MSE between the original image and the image reconstructed from the sequence of fixations. The MSE serves as a measure of total uncertainty about the original image. We find that the CEM algorithm also does quite well at reducing total uncertainty in individual

images: The MSE values average about 10% higher than optimal (see Figs. 3.7C and 3.7D).

Although the CEM algorithm is quite simple and is based only on contrast statistics, it performs remarkably well at reducing total image uncertainty, and hence it may be of practical value in certain surveillance and robotic applications involving foveated imaging. For example, if there is time to make only a few fixations with a remote robotic or surveillance camera, then the CEM algorithm could be used to select those few fixations, assuming the goal is to reconstruct the image as accurately as possible. The algorithm is amenable to parallel computing and runs at a respectable speed (a fixation every couple of seconds) on a standard personal computer.

The performance of the CEM in predicting human fixation patterns will be described in detail in Chapter 7. As explained in Chapter 1, visual fixation patterns are very much dependent on the high-level visual tasks performed by the organism. In reading, saccade lengths tend to be short and the fixation patterns stereotypical because, for the most part, words must be read sequentially for the communication to be understood [110]. In search tasks where the observer is trying to find a specific target or class of targets, saccade lengths tend to be longer and the fixation patterns more random than in reading because the eye is drawn to any likely target location in the image [117-118]. In general, human fixation patterns are probably different for every kind of perceptual or cognitive task that is performed [37].

The class of tasks that we are concerned with in this dissertation is visual search tasks—and in particular, tasks in which the objective is to acquire as much information about the image as possible. In such tasks, the goal is to learn as much as possible about a

scene in a few fixations, so that the scene can be distinguished from other scenes at a later time. Picking fixations that minimize contrast entropy is a relatively simple and efficient way to gain information about the scene because the fixation selection requires no encoding of spatial structure, no pattern recognition, and little other high-level processing. Minimizing contrast entropy involves only encoding local contrasts and pooling them in a way that is weighted by the eccentricity and the contrast. This is the kind of processing that could be done in a fairly low level and automatic way, without placing great demands on high-level processes that require more attentional resources. What makes minimizing contrast entropy particularly appealing for this class of task is that it also does a good job of reducing total uncertainty about the image. Thus, selecting fixations by minimizing contrast entropy will, to good approximation, maximize the amount of image structure available to the cortex for extraction and storage in memory. The fact that the CEM makes detailed predictions about the fixation patterns enables us to test the CEM algorithm for visual search tasks which as we demonstrate in Chapter 7, performs very well in matching human fixation patterns for visual search tasks.

An important aim of this study was to measure the contrast statistics of natural images for foveated visual systems. We have focused on the relevance of these statistics for fixation selection, but it is obvious that they must be of at least some relevance for many tasks that involve information integration or comparison across the visual field. The fact that the posterior probability distribution of the true unblurred local contrast is characterized by very simple formulas should make it possible to incorporate these natural scene statistics into various Bayesian models of perceptual performance.

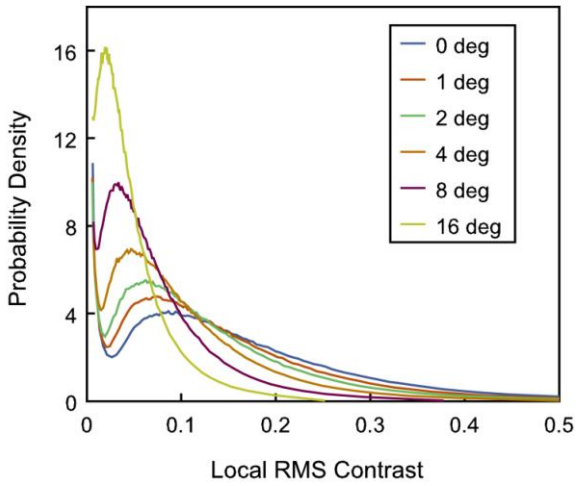


Fig. 3.1: Probability distributions of local rms contrast for various levels of blur based on the human contrast sensitivity function at different retinal eccentricities. These distributions were obtained by randomly sampling small patches from 300 calibrated natural images.

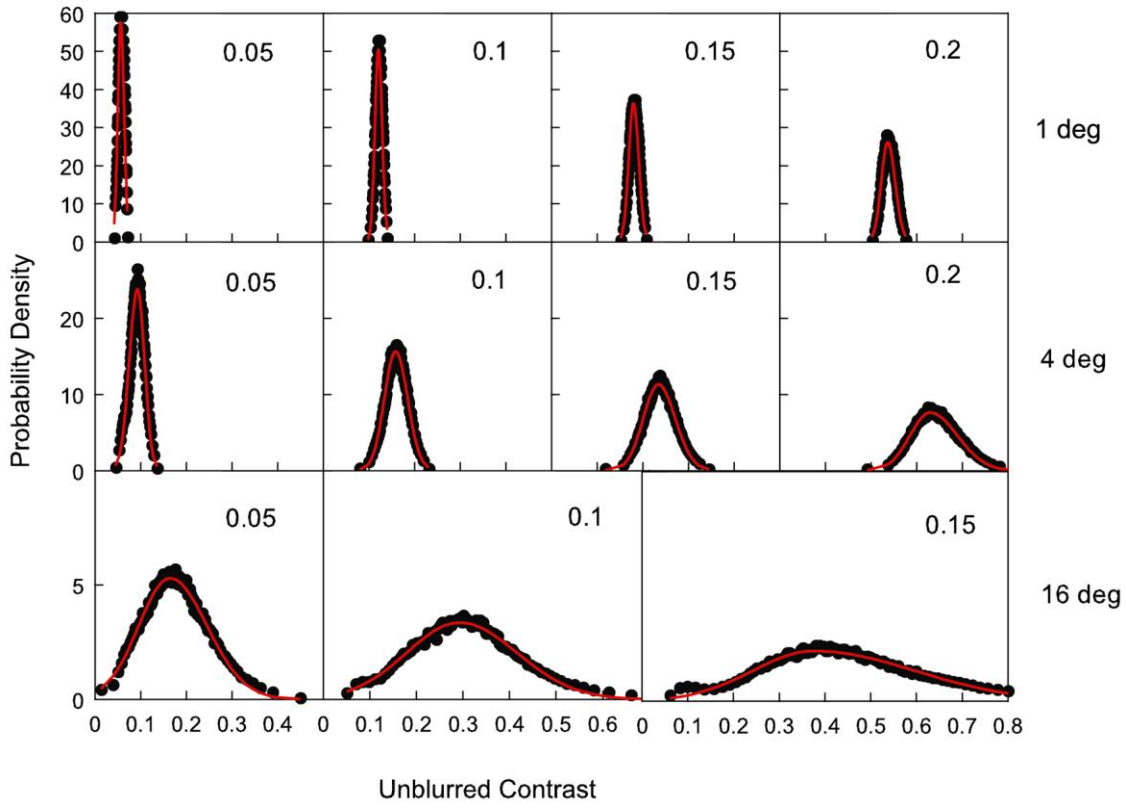


Fig. 3.2: These plots show examples of the conditional probability distributions of local rms contrast in unblurred images, given the local rms contrast in the blurred versions of the images (columns) and given the retinal eccentricity (rows). The solid symbols are empirical histograms computed from 300 natural images that contained no man-made objects. The smooth curves are the bestfitting skewed Gaussian distribution (a Gaussian with different standard deviations above and below the mode).

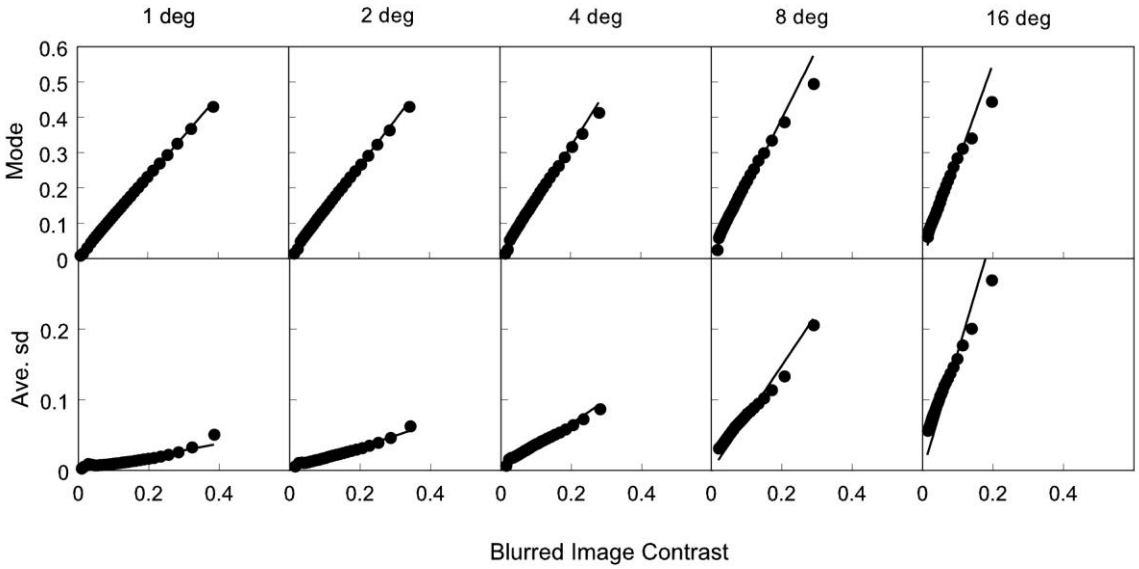


Fig. 3.3: Modes and average standard deviations of the conditional probability densities are plotted as a function of blurred image contrast and retinal eccentricity. The average standard deviation is the average of the two standard deviation parameters in the skewed Gaussian distribution. See Fig. 3.2 for examples of the conditional densities and fits of the skewed Gaussian distribution. The curves are best fitting straight lines through the origin.

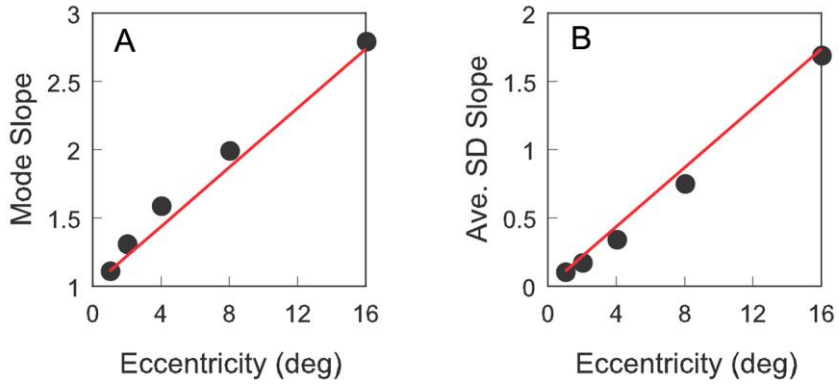


Fig 3.4: Slopes of the linear functions in Fig. 3.3. A, Slope of the contrast versus mode plot as a function of retinal eccentricity. B, Slope of the contrast versus average standard deviation plot as a function of retinal eccentricity. The curves show the predictions of the linear model: $\hat{c} = k\varepsilon c + c$ and $\bar{\sigma} = k\varepsilon c$, where k=0.105

Fig 3.5: Images used to test a fixation selection algorithm based on the principle of minimizing contrast entropy.

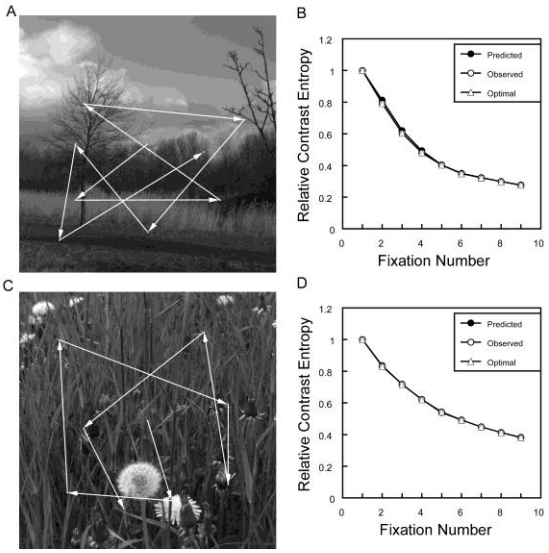


Fig. 3.6: Fixation points selected by the principle of minimizing total contrast entropy (contrast uncertainty), using the average local contrast statistics of natural images. A, Sequence of nine fixations (eight saccades) for a distant image containing sky, ground, and trees. B, Relative contrast entropy as a function of fixation number for the image in A (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). C, Sequence of nine fixations (eight saccades) for a close-up image containing foliage. D, Same type of plot shown in B.

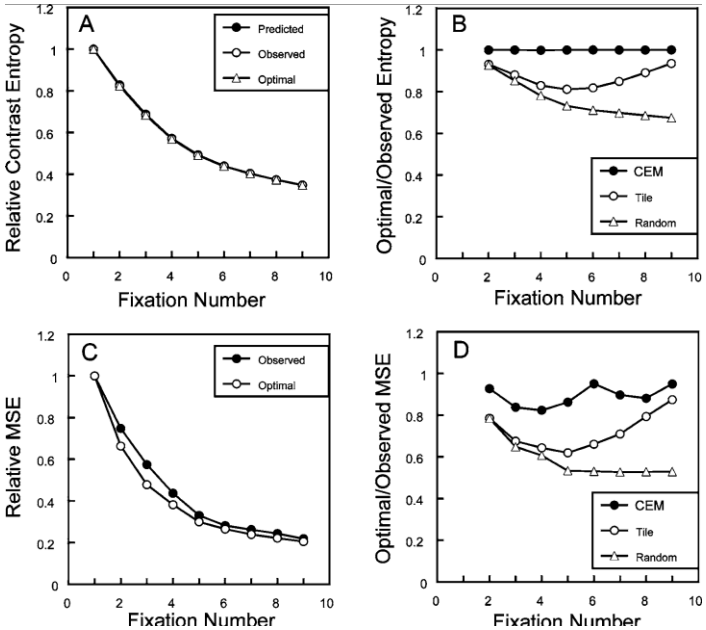


Fig 3.7: Average fixation selection performance for the 16 test images in Fig. 3.5. A, Relative contrast entropy as a function of fixation number (open circles), predicted relative contrast entropy before the fixation was made (solid circles), and optimal relative contrast entropy that could be obtained (open triangles). B, Ratio of the optimal contrast entropy that could be obtained to the contrast entropy that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles). C, Relative mean squared error (MSE) between the original (unblurred) image and the image reconstructed from the fixations up to and including the fixation number given on the horizontal axis: CEM algorithm (solid circles), optimal (open circles). D, Ratio of optimal MSE that could be obtained to the MSE that was obtained: CEM algorithm (solid circles), tiling algorithm (open circles), random algorithm (open triangles).

Chapter 4

MICA: A Multilinear ICA Decomposition for Natural Image Modeling

4.1 Introduction

In the previous chapter we studied, in detail, a simple formulation of the contrast statistics of natural images which has a direct application to deriving optimal contrast-based fixations. The next major goal of our dissertation is to devise optimal texture-based fixation strategies of natural images. Texture, unlike contrast, is a region based concept—the characterization of which requires probabilistic descriptions on multi-dimensional spaces. Thus we are naturally led to the task of probabilistically examining the spatial structure of natural images in more detail—and in a way that can lead to useful formulations of the problem of optimally extracting textural information from natural scenes. This chapter examines a powerful and elegant new approach of capturing the spatial image statistics of natural images.

The construction of accurate prior models of natural image source data is essential to many applications (such as low-level vision) for which unsupervised learning methods must be applied due to the inherent lack of labeled training sets. Such prior models give a framework in which to correctly interpret the data, thereby serving as the basis for subsequent analysis viewed from different levels of abstraction. There are a variety of classical unsupervised methods that exist for this purpose, including Principle

Component Analysis (PCA), Independent Component Analysis (ICA) and Multidimensional Scaling (MDS) [17].

Among these classic tools, ICA has several important and distinguishing characteristics. Denote the probability of the source that we are modeling by $P(X)$, where X is a random vector whose realizations have dimensionality d. The goal of ICA is to factor the probability distribution of the source into a product of distributions: $P(X) = \prod_{i=1}^{d} p(s_i)$, where $\{s_i = X * \phi_i\}_{i=1}^{d}$ are filtered responses of the source. The filters $\{\phi_i\}_{i=1}^{d}$ are the ICA filters of the source. Statistical algorithms for computing the ICA filters have been the subject of intense study over the past decade [119], most of which involve the construction of different cost functions (usually variations of the maximum likelihood cost function).

The independent directions that emerge from an ICA decomposition can be fruitfully utilized by reducing the d-dimensional problem into d independent 1-D problems. Furthermore, ICA decompositions of data having heavy tailed marginals (as is for example observed in NSS applications) tend to favor sparse representations [120]. Sparse representations are useful for many applications that seek to efficiently represent and process the data.

However, in spite of these potential advantages, in reality the statistics of most real-world sources, such as natural image patches, cannot be strictly factored into a simple product. As a result, the so-called independent components contain significant mutual dependencies between them [121].

In order to ameliorate this situation we propose a refinement of the classic ICA model, which we call the Multilinear ICA (MICA) decomposition, wherein the

dependencies between pseudo-independent components are captured using a multilinear representation of P(X):

$$P(X) = \frac{1}{Z} g(J) \prod_{i=1}^{d} p(s_i)$$

where $g : J = [s_1, ..., s_d] \rightarrow R$ and $Z \in R$ is a normalizing constant. Of all possible multilinear expansions of this form that could describe the source distribution, we seek the one that makes the representation of the source as sparse as possible, i.e., which minimizes the contribution of $g(J)$. Naturally we are interested in closed form approximations for such a $g(J)$. The multilinear form thus obtained retains all the attractive properties of the ICA decomposition, and at the same time lumps the interactions of the filtered responses into the function $g(J)$. Of course, when $g(J)$ is separable with respect to the filter responses, this reduces to the classical ICA representation.

The success of our proposed method depends upon the accuracy of the numerical approximation of $g(J)$. Analytical methods of approximating $g(J)$ using Taylor expansions seem formidable. Further it is necessary to estimate Z which, in general, requires tedious Monte Carlo simulations.

In Section 4.2, we introduce a non-linear system model that enables us to circumvent the above issues. We call the resulting refinement of ICA the *Multilinear ICA* (*MICA*) *Model*. We successfully deploy the new method to model natural scene textures in Section 4.3, and demonstrate advantages relative to classical ICA. We conclude in Section 4.4 with a discussion of possible applications of the MICA model together with some open problems.

4.2 The Multilinear ICA Model

A. Overview and Parameter Description

Consider the classical ICA model where the observation vector is modeled: $\tilde{s} = Bz$; where $\tilde{s} = [\tilde{s}_1,...,\tilde{s}_d]^T \in R^d$, $z = [z_1,..., z_d]^T \in R^d$, d is the intrinsic dimensionality of the data, and $B \in R^{dxd}$ is a full-rank matrix. The goal of ICA is to find a matrix B such that the resulting components of *z* are independent random variables.

However, for many real-world sources, such as natural images, such an ideal decomposition is not possible and so the components of *z* will contain residual dependencies. Our aim is to explicitly capture these dependencies. In doing so we must first recognize that *z* cannot be further decomposed as a combination of independent sources via another full-rank matrix! It is possible, however, that *z* can be decomposed with respect to an under-complete linear model, but this requires knowledge of the subspace dimensionality.

An alternate view which we explore in this chapter is that, given knowledge of the intrinsic dimensionality *d,* the residual dependencies can be captured via non-linear combinations of independent sources. The choice of the non-linearity, as well as of the source distribution, must be as simple as possible, and yet must successfully account for the probabilistic structure of the observed natural image sources. To simplify matters further, we first concern ourselves only with modeling unimodal distributions which, as shown in Section 4.3, appears to be well-suited to many natural image textures. Later on we suggest how to extend this to multimodal cases via mixtures of MICA models. We first focus on the complete basis case (i.e. where *B* is a full-rank matrix). Later, we will demonstrate how these ideas can be extended to the under-complete case in a straightforward manner.

Perhaps the simplest non-linear system that one can hypothesize for natural image source modeling is a quadratic channel. In our experiments with natural image textures, we found that the hybrid linear-quadratic model (stimulated by *i.i.d.* Gaussian sources) shown in Fig. 4.1 can successfully account for the probabilistic structure of natural image patches. We now describe this non-linear system in detail.

The observable image source data that we are modeling is $\tilde{s} \in R^d$. $B \in R^{dxd}$ is a full-rank matrix initially chosen as the matrix associated with the classical ICA decomposition of $\tilde{s}$ which will be re-estimated in subsequent iterations. The system *F* in Fig. 4.1 models the residual interaction between the components of $z \in R^d$. It consists of a core non-linearity φ preceded by a linear system $y = As + \gamma$, where $y = [y_1,..., y_d]^T \in R^d$, $\gamma = [\gamma_1,...,\gamma_d]^T \in R^d$, and $s = [s_1,..., s_d]^T \in R^d$ are *i.i.d.* Gaussian: $s_i \sim \aleph(0,1)$. The density of the i^{th} Gaussian channel is denoted $q(s_i)$. The Gaussian channel variances are $\sigma = [\sigma_1,..., \sigma_d]^T \in R^d$, $\mu = [\mu_1,..., \mu_d]^T \in R^d$ is an additive mean adjusting vector, and $\beta = [\beta_1,..., \beta_d] \in R^d$ is a multiplicative vector that is applied (component-wise) to all channels, and which determines the effective non-linearity of the channels. Finally, $C = [C_{i,j}] = [C_1^T,...,C_d^T] = A^{-1} \in R^{dxd}$ is an invertible linear transformation of the *i.i.d.* Gaussian sources that determines the interaction of the Gaussian sources.

B. Structure of the MICA Distribution

The non-linearity φ consists of complementary linear and quadratic channels. Operators u_1 and u_2 are complementary limiters: $u_1(y_i) + u_2(y_i) = 1$, $u_1(y_i), u_2(y_i) \geq 0$ for $1 < i < d$.

A simple choice of limiters which we have found to be useful for modeling natural image textures (see Section 4.3), are the complementary step functions:

$$u_1(y_i) = u(y_i + 1) - u(y_i - 1),\ u_2(y_i) = 1 - u_1(y_i)$$

where *u(x)* is the unit step function. From this we obtain:

$$\varphi(y) = y u_1(y) + \varphi_q(y) u_2(y)$$

where $\varphi_q(y) = y^2 \operatorname{sgn}(y)$ (throughout, operations on y are applied component-wise). The function φ is plotted in Fig. 4.11.

For this choice of (u_1, u_2):

$$\tilde{\varphi}[\beta(z-\mu)] \equiv \varphi^{-1}[\beta(z-\mu)] = \beta(z-\mu) u_1[\beta(z-\mu)] + \varphi_q^{-1}[\beta(z-\mu)] u_2[\beta(z-\mu)]$$

where

$$\varphi_q^{-1}[\beta(z-\mu)] = \left\{ \sqrt{|\beta_i(z_i - \mu_i)|} \operatorname{sgn}[\beta_i(z_i - \mu_i)] \right\}_{i=1}^{d}.$$

Since the non-linearity is invertible, system F is also:

$$s = F^{-1}[(z-\mu)] = C\{\tilde{\varphi}[\beta(z-\mu)] - \gamma\}.$$

The distribution of $\tilde{s}$ then has the following form (where throughout $|\cdot|$ is the matrix determinant):

$$P(\tilde{s}) = \frac{1}{|B|} p(z) = \frac{1}{|B|} \cdot \frac{1}{|J(F)|} \prod_{k=1}^{d} q\left\{ F_k^{-1}[\beta(z-\mu)] \right\}$$

where $z = B^{-1}\tilde{s}$ and $q(s_k)$ is the k^{th} Gaussian source channel. Expanding *p(z)* yields the MICA model:

$$p(z) = \frac{K}{|J(F)|} g(J) \prod_{k=1}^{d} p((z_i - \mu_i)\beta_i) \tag{4.1}$$

where

$p[(z_i-\mu_i)\beta_i] = K_i \exp\left(-a_i\left[\tilde{\varphi}[\beta_i(z_i-\mu_i)]-c_i\right]^2\right)$, K_i is a normalizing constant, $a_i = \frac{1}{2}\sum_{k=1}^{d}\frac{C_{k,i}^2}{\sigma_k^2}$, and $c_i = \gamma_i + \frac{\sum_{j\neq i}\sum_{k=1}^{d}\frac{C_{k,j}C_{k,i}}{\sigma_k^2}\gamma_j}{\sum_{k=1}^{d}\frac{C_{k,i}^2}{\sigma_k^2}}$.

Also $g(J) = \exp\left\{-\sum_{i\neq j} G_{i,j}\varphi[\beta_i(s_i-\mu_i)]\varphi[\beta_j(s_j-\mu_j)]\right\}$, where $K = \frac{\exp\left[-\sum_{k=1}^{d}\frac{(C_k^T\gamma)^2}{2\sigma_k^2}\right]}{(2\pi)^{d/2}\prod_{k=1}^{d}\sigma_k K_k}$, and $G_{i,j} = -\sum_{k=1}^{d}\frac{C_{k,i}C_{k,j}}{\sigma_k^2}$.

In (4.1), $J(F)$ is the Jacobian of the transformation F, for which the following theorem (proved in the Appendix) yields a closed form expression.

Theorem 4.1: The Jacobian of the transformation F is:

$$J(F) = \frac{1}{|C|}\prod_{k=1}^{d}\frac{\psi[\beta_k(z_k-\mu_k)]}{\beta_k},$$

where, $\psi[\beta_k(z_k-\mu_k)] = u_1[\beta(z_k-\mu_k)] + 2\,|\,\varphi_q^{-1}[\beta_k(z_k-\mu_k)]\,|\,u_2[\beta_k(z_k-\mu_k)]$ ♣

The *MICA interaction matrix* $[G_{i,j}]$ captures interactions between the MICA components. In particular, when $[G_{i,j}]_{i\neq j} = 0$, the MICA components are independent. Fig. 4.9(a-d) shows the frequency response of a few of the MICA filters (i.e. derived from the *B*-matrix in Fig.4.1) of the Gravel texture as described in more detail in Section 4.3. We observe that these MICA filters exhibit bandpass like behavior and that, in general, there will be overlap among the spectra between the various MICA components. The overlapping of spectra, however, does not by itself indicate the degree of dependence

between the MICA filters. The latter is captured more accurately by the MICA interaction matrix which, for the Gravel texture, is shown in Fig. 4.9(e). In particular, the greater the value of $G_{i,j}$, the greater is the degree of statistical dependency between the corresponding MICA filters.

The parameter β determines the degree of nonlinearity in the system which can be qualitatively understood as follows: when training the MICA model (given the filtered data *z*), β determines the extent to which *w* is scaled inside the unit interval and consequently determines (after σ is adjusted as a part of the MICA optimization) the extent to which the linear channel of the system is active. Thus β determines the tradeoff between the linear and quadratic models when determining the optimal tail and peak behaviors of the MICA distribution. Once the above parameters have been adjusted, the vector μ is chosen to optimally adjust the mean of MICA. Finally, γ determines the skew of the marginal distributions by asymmetrically assigning the non-linearity within the effective domain of the distribution.

C. Parameter Estimation for the MICA Model

We estimate the optimal parameters of the MICA model (4.1) by employing a steepest gradient algorithm with respect to the log-likelihood function:

$$\log[p(z)] = \log\left[\frac{\mathrm{abs}(|C|)}{(2\pi)^{d/2}}\prod_{k=1}^{d}\left(\frac{\beta_k}{\sigma_k}\right)\right] - \log\left(\prod_{i=1}^{d}\left|\psi[\beta_i(z_i-\mu_i)]\right|\right) - \sum_{k=1}^{d}\frac{(C_k^T\gamma)^2}{2\sigma_k^2}$$

$$+\frac{1}{2}\sum_{i=1}^{d}\left(\sum_{k=1}^{d}\frac{C_{k,i}^2}{\sigma_k^2}\right)\left\{\tilde{\varphi}[\beta_i(z_i-\mu_i)]\right\}^2 - \sum_{k=1}^{d}\left(\sum_{k=1}^{d}\frac{C_k^T\gamma}{\sigma_k^2}C_{k,i}\right)\tilde{\varphi}[\beta(z_i-\mu_i)]$$

$$-\sum_{i\neq j}G_{i,j}\tilde{\phi}[\beta(z_i-\mu_i)]\tilde{\phi}\left[\beta(z_j-\mu_j)\right] \qquad (4.2)$$

From (4.2), the gradient of the log-likelihood function with respect to the different parameters can be computed in a straightforward manner:

$$\frac{\partial \log[p(z)]}{\partial C_{m,n}} = \frac{(-1)^{m+n}|C^{m,n}|}{\text{abs}(|C|)}\text{sgn}|C| - \left(\frac{C_m^T\gamma}{\sigma_m^2}\right)\gamma_n - \frac{C_{m,n}}{\sigma_m^2}\left\{\tilde{\varphi}[\beta_n(z_n-\mu_n)]\right\}^2$$

$$+\tilde{\varphi}[\beta_n(z_n-\mu_n)]\left[\frac{C_m^T\gamma}{\sigma_m^2}+\frac{C_{m,n}\gamma_n}{\sigma_m^2}\right] - \sum_{i\neq n}\frac{C_{m,i}}{\sigma_m^2}\tilde{\varphi}[\beta_i(z_i-\mu_i)]\tilde{\varphi}[\beta_n(z_n-\mu_n)] \quad (4.3)$$

$$\frac{\partial \log[p(z)]}{\partial \sigma_m} = -\frac{1}{\sigma_m} + \frac{(C_m^T\gamma)^2}{\sigma_m^3} + \frac{1}{\sigma_m^3}\sum_{i=1}^{d} C_{m,i}^2\left\{\tilde{\varphi}[\beta_i(z_i-\mu_i)]\right\}^2 - \frac{2}{\sigma_m^3}\sum_{i=1}^{d}(C_m^T\gamma)C_{m,i}\tilde{\varphi}[\beta_i(z_i-\mu_i)]$$

$$+\frac{2}{\sigma_m^3}\sum_{i\neq j} C_{m,i}C_{m,j}\tilde{\varphi}[\beta_i(z_i-\mu_i)]\tilde{\varphi}[\beta_j(z_j-\mu_j)] \quad (4.4)$$

$$\frac{\partial \log(p(z))}{\partial \gamma_m} = -\sum_{k=1}^{d}\frac{C_k^T\gamma}{\sigma_k^2}C_{k,m} + \sum_{i=1}^{d}\tilde{\varphi}[\beta_i(z_i-\mu_i)]\left(\sum_{k=1}^{d}\frac{C_{k,m}C_{k,i}}{\sigma_k^2}\right) \quad (4.5)$$

$$\frac{\partial \log[p(z)]}{\partial \mu_m} = \frac{\beta_m u[\beta_m(z_m-\mu_m)-1]}{\varphi_q^{-1}[\beta_m(z_m-\mu_m)]}\cdot\frac{1}{|\psi[\beta_m(z_m-\mu_m)]|}\left[\sum_{k=1}^{d}\frac{C_{k,m}^2}{\sigma_k^2}\tilde{\varphi}[\beta_m(z_m-\mu_m)]\right.$$

$$\left.+\sum_{i\neq m}\left(\sum_{k=1}^{d}\frac{C_{k,i}C_{k,m}}{\sigma_k^2}\right)\tilde{\varphi}[\beta_m(z_m-\mu_m)] - \sum_{k=1}^{d}\frac{C_k^T\gamma}{\sigma_k^2}C_{k,m}\right] \quad (4.6)$$

$$\frac{\partial \log[p(z)]}{\partial \beta_i} = \frac{d}{\beta_i} - \left(\frac{\chi_i}{\prod_{i=1}^{d}|\psi[\beta_i(z_i-\mu_i)]|}\right) - \sum_{i=1}^{d}\left(\sum_{k=1}^{d}\frac{C_{k,i}^2}{\sigma_k^2}\right)\tilde{\varphi}[\beta_i(z_i-\mu_i)]\tilde{\varphi}'[\beta_i(z_i-\mu_i)]z_i$$

$$-\sum_{i=1}^{d}\left(\sum_{k=1}^{d}\frac{C_k^T\gamma}{\sigma_k^2}C_{k,i}\right)\tilde{\varphi}'[\beta_i(z_i-\mu_i)](z_i-\mu_i)$$

$$-\sum_{i\neq j}\left(\sum_{k=1}^{d}\frac{C_{k,i}C_{k,j}}{\sigma_k^2}\right)\left\{\tilde{\varphi}[\beta_i(z_i-\mu_i)]\tilde{\varphi}'\left[\beta_j(z_j-\mu_j)\right](z_j-\mu_j)\right.$$

$$\left.+\tilde{\varphi}\left[\beta_j(z_j-\mu_j)\right]\tilde{\varphi}'[\beta_i(z_i-\mu_i)](z_i-\mu_i)\right\} \tag{4.7}$$

where

$$\chi_i=\beta_i\sum_{i=1}^{d}\left\{\prod_{j\neq i}^{d}|\psi[\beta_i(z_i-\mu_i)]|\right\}.\left(u[\beta_i(z_i-\mu_i)-1]+\frac{u_2[\beta_i(z_i-\mu_i)]}{\varphi_q^{-1}[\beta_i(z_i-\mu_i)]}\right)z_i$$

$C^{m,n}$ is the co-factor matrix of C with respect to (m, n), and $C_{m,n}$ is the $(m, n)^{th}$ entry of C.

Our goal is to obtain a multilinear expansion of $P(\tilde{s})$ corresponding to a sparse representation of the source. This can be accomplished by initializing B with the matrix associated with the classical ICA decomposition of source $\tilde{s}$. Then $z=B^{-1}\tilde{s}$. A gradient descent algorithm then obtains the optimum parameters $C=A^{-1}$, σ, γ, μ and β using the above expressions. A multilinear expansion of $P(\tilde{s})$ is obtained as in (4.1), the structure of which is specified by these parameters. The estimate of B can be further refined by fixing the parameters, then invoking a gradient descent algorithm. Let $D=B^{-1}$; then $p(\tilde{s})=\text{abs}(|D|)p(z)$. The gradient of $\log[p(\tilde{s})]$ with respect to $D_{m,n}$ (the $(m, n)^{th}$ entry of D) is:

$$\frac{\partial\log[p(\tilde{s})]}{\partial D_{m,n}}=\frac{(-1)^{m+n}|D^{m,n}|\,\text{sgn}(|D|)}{\text{abs}(|D|)}-\frac{\beta_m\left\{u[\beta_m(z_m-\mu_m)-1]+\dfrac{u_2[\beta_m(z_m-\mu_m)]}{\varphi_q^{-1}[\beta_m(z_m-\mu_m)]}\right\}\tilde{s}_n}{\left|\psi[\beta_m(z_m-\mu_m)]\right|}$$

$$-\left(\sum_{k=1}^{d}\frac{C_{k,m}^2}{\sigma_k^2}\right)\tilde{\varphi}[\beta_m(z_m-\mu_m)]\tilde{\varphi}'[\beta_m(z_m-\mu_m)]\tilde{s}_n$$

$$-\sum_{j\neq m}\left(\sum_{k=1}^{d}\frac{C_{k,m}C_{k,j}}{\sigma_k^2}\right)\tilde{\varphi}\left[\beta_j(z_j-\mu_j)\right]\tilde{\varphi}'[\beta_m(z_m-\mu_m)]\tilde{s}_n$$

$$-\left(\sum_{k=1}^{d}\frac{C_k^T\gamma}{\sigma_k^2}C_{k,m}\right)\tilde{\varphi}'\left[\beta_m(z_m-\mu_m)\right]\tilde{s}_n \tag{4.8}$$

Once B is computed, the two-step process of estimating $(C,\sigma,\gamma,\mu,\beta)$ followed by re-estimating B may be performed until a desired level of accuracy is achieved. However, in our simulations we find that a single estimate of $(C,\sigma,\gamma,\mu,\beta)$, without subsequent re-estimation of B generally outperforms classical ICA modeling on natural image textures (using the Kullback-Leibler divergence as a measure of performance) as shown in Section 4.3 below. The high-level MICA algorithm thus described is summarized in Fig. 4.10.

We have found it convenient to heuristically estimate β instead of employing (4.7). Our heuristic for estimating β is described and motivated as follows. Consider the case where the distribution of the i^{th} data channel is heavy-tailed (high-kurtosis). Modeling the i^{th} channel histogram by a Laplacian distribution

$$f(z\,|\,b_i)=\frac{1}{2b_i}\exp\left(-\frac{|\,z-\mu_i\,|}{b_i}\right),$$

the parameter b_i can be estimated from the data using the following closed form expression [122]:

$$\hat{b}_i=\frac{1}{N}\sum_{j=1}^{N}|\,z_j^i-\hat{\mu}_i\,|$$

where $\hat{\mu}_i$ is the sample median of the i^{th} channel data $\{z_j^i\}_{j=1}^N$. From (4.1), β_i can be thought of as proportional to $1/b_i$. This yields a heuristic for the initial estimate of β_i: $\beta_i^0=1/(\hat{b}_i)$. Further refinements can be obtained in subsequent iterations by observing that in (4.1), a more accurate relationship between the b_i and β_i is as follows: $\beta_i\approx 1/(\hat{b}_i a_i)$. The

k^{th} estimate of β_i is $\beta_i^k \approx 1/(\hat{b}_i a_i^{k-1})$, where a_i^{k-1} is the $(k\text{-}1)^{st}$ estimate of a_i obtained when using β_i^{k-1}. As shown in Section 4.3, the initial estimate $\beta_{high-kurt} = \beta_i^0$ yields better performance than ICA, even without subsequent re-estimation of β_i.

For the case where the i^{th} data channel is not heavy tailed (i.e. the low-kurtosis case), it is intuitive to emphasize the linear part of $\tilde{\varphi}$—tantamount to initializing β_i such that $\beta_i \le \frac{1}{\max_n [z_i(n)]}$. In practice, $\beta_{low-kurt} = \beta_i^0 = \frac{1}{\max_n [z_i(n)]}$ suffices for low-kurtosis cases. As with the high kurtosis case, we find it unnecessary to update the estimate of β_i at every iteration, but instead use the initial estimate β_i^0 throughout the optimization process.

To simplify matters further, we employ a single scalar parameter β that we apply to *all* channels. To estimate β we employ a similar heuristic as above. For high-kurtosis, use $\beta_{high-kurt} = 1/\hat{b}$ where $\hat{b} = \frac{1}{Nd}\sum_{i,j} | z_j^i - \hat{m}_i |$ and $\hat{m}_i$ is the sample median of $\{z_j^i\}_{i,j}$. Similarly, for low kurtosis take $\beta_{low-kurt} = \frac{1}{\max_{i,j} [z_i(j)]}$.

As a simple way of deciding the β estimate to be used, we measure the local sample kurtosis κ. The high-kurtosis heuristic is used when $\kappa \ge 4$, and the low-kurtosis heuristic when $\kappa < 4$.

D. Extension to Under-Complete MICA models

Consider the case where the observation $\tilde{s}$ is modeled as follows: $\tilde{s} = Bz$, where $\tilde{s} = [\tilde{s}_1, ..., \tilde{s}_l]^T \in R^l$ and $z = [z_1, ..., z_d]^T \in R^d$, such that $d < l$, i.e. $B \in R^{lxd}$ is an under-complete matrix. Under-complete models arise in situations where dimensionality reduction is

required in order to model the data in an appropriate subspace.

As before, we ask whether multilinear modeling can accurately capture the statistical dependencies between the components of z. As shown in Section 4.3, the answer to this is affirmative. A simple way of assessing the performance of MICA for under-complete models is to first consider the corresponding complete basis case where $\tilde{s} = \tilde{B}\tilde{z}$, $\tilde{s} = [\tilde{s}_1,...,\tilde{s}_l]^T \in R^l$ is the observed source vector as before, $\tilde{B} \in R^{lxl}$ and $\tilde{z} = [\tilde{z}_1,...,\tilde{z}_l]^T \in R^l$. The matrix $\tilde{B}$ is initialized with the classic ICA matrix as described in previous sections. Now we let z constitute the d most significant ICA components of $\tilde{z}$: $z = \tilde{z}(1:d) = V\tilde{s}$, where $V = \tilde{B}^{-1}(1:d,:)$ (assuming that the rows of $\tilde{B}$ are arranged according to the energy corresponding to the corresponding directions in the data space). We now model the components of z by the complete MICA model developed in previous sections. In order to evaluate the performance of MICA, first obtain estimates of the original source vector $\tilde{s}$ by assigning the initial estimate of matrix B to be the pseudo-inverse of V, i.e. $B = (VV^*)^{-1}V$. As in previous sections, we do not re-estimate B but just use the initial estimate along with the optimal MICA parameters computed as above.

4.3 Simulation Results

We define the M x M *image patch statistics* of an N x N image region to be the joint distribution of the random variables (pixel values) from M x M patches that sample the image region. In this chapter we are specifically interested in modeling the M x M image patch statistics of natural scenes. We first demonstrate the performance of MICA for the case of complete basis for M = 3. Since, for larger patch sizes the MICA optimization algorithm becomes more computationally cumbersome, we demonstrate how under-

complete MICA models can be successfully exploited to reduce complexity.

To evaluate the complete basis case, we uniformly sampled texture images obtained from the USC-SIPI Brodatz database [123] with N_{ptch} = 2000 patches of size M x M. An ICA was then performed on the data vectors obtained from each texture using Comon's algorithm [124] to obtain the matrix B. Subsequently the parameters $(C, \sigma, \gamma, \mu, \beta)$ of the MICA model were estimated as described in Section 4.2. The parameter β, as mentioned earlier, was estimated heuristically at the outset of the simulation and held to a constant value throughout. To limit computation time, the optimization routine for estimating (C, σ, γ, μ) was forced to terminate after only a few iterations.

Parameter initialization prior to running the optimization routine was performed as follows. Matrix C was initialized to the identity matrix, the entries of σ were initialized to 0.5, and the entries of μ were adjusted such that each of the z channels are zero-mean. After running the MICA optimization routine, setting the γ parameter to the skew of the corresponding data channels gives consistently good performance. The intuitive reason for this choice of γ can be seen by considering the generative model in Fig. 4.1. Once all the parameters of the MICA model have been adjusted, varying γ determines the asymmetry with which samples are exposed to the linear and quadratic channels: by varying γ, we can directly control the skew of the resulting distribution.

Thereafter, for each texture, we compared the data distribution of each channel derived from test data sets (different from the training data sets) to the corresponding distribution predicted by the ICA and MICA models. In addition, the average of all the data channels was also compared with that predicted by the ICA and MICA models. Simulation of the MICA model was accomplished by generating *d i.i.d.* zero mean, unit-

variance Gaussian channels as shown in Fig. 4.1, and plotting the histograms outputs of the channels when the optimal parameters (for the texture being modeled) were used. The ICA model was simulated by first computing the empirical distributions of each channel, which were then independently sampled and processed by the matrix B. The histograms of the ICA and MICA channels were then compared with the corresponding channels of the original data distributions using the Kullback-Leibler divergence (KLD) [116]. The above procedure was repeated over several trials, and the average KLD for each channel (for both ICA and MICA) computed with respect to the corresponding channels of the data distributions.

Figures 4.2(a)-4.8(a) depict texture images taken from the Brodatz database [123]. Figs. 4.2(b)-4.8(b) show the histograms of two of the channels corresponding to each of the textures, as well as both the corresponding computed ICA distributions and the corresponding computed MICA distributions. Also shown are the histograms of the data distributions when all of the data channels of the corresponding textures are averaged together as well as the corresponding computed ICA and MICA distributions. The heuristic strategy used to compute the parameter β for each of these cases is also indicated. Ideally, of course, one can compute β using the optimal derivation given earlier, but the heuristic kurtosis-based approach has proven to yield efficient, near-optimal MICA solutions.

In Figs. 4.2-4.4, it is apparent that the MICA model allows for significantly improved approximation of the original data distributions as compared to the classic ICA model. In Figs. 4.5-4.8, it is apparent that MICA does a better job in capturing the kurtosis of the channels and does a slightly better job in capturing the skew of the original data

distributions; as for example in Figs. 4.6-4.7. Furthermore in all cases, there is improved approximation of the peak and tail behavior of the original data distributions as compared to ICA. Quantitative evaluation of the MICA model for the complete basis case is provided in Table 4.1, where the relative improvement of MICA relative to classic ICA is measured as:

$$\theta_{KLD}^{MICA} = \frac{KLD(ICA) - KLD(MICA)}{KLD(ICA)}$$

where $KLD(MICA)$ is the KLD between the MICA channels and the corresponding original data distributions (averaged across all channels), $KLD(ICA)$ is the corresponding average KLD for the ICA model, and θ_{KLD}^{MICA} is the relative improvement due to MICA over classic ICA with respect to KLD. It is apparent from Table 1 that the relative performance of MICA is consistently better than that of classic ICA for all natural scene textures.

Similarly, Table 4.2 quantifies the performance of the under-complete MICA model for ($M = 5$, $d = 9$). The initialization of the parameters before the optimization is similar to that described before, except that the entries of μ are set to zero without subsequent setting of γ to the skew of the data channels. Furthermore, β is always chosen according to the low-kurtosis heuristic. Unlike the complete basis case, $p(z)$ and $p(\tilde{s})$ are no longer related by a simple scale factor, and so the roles of the different parameters in determining $p(\tilde{s})$ becomes more complicated. Nevertheless we find, as shown in Table 4.2, that MICA consistently outperforms classical ICA using our simple approach. Under-complete models are useful when is desired to use large patch sizes to sample the image, yet make the problem computationally tractable by working in a lower

dimensional sub-space. These results demonstrate that the basic idea of multilinear modeling of probability distributions can be successfully extended to under-complete cases.

We further point out that a comprehensive approach to finding the optimal MICA model parameters would be to incorporate an additional simulation optimization phase where the Gaussian random vector s is generated to drive the optimization of the parameters to match the desired data distributions. Such a procedure is likely to be more efficient than a Monte-Carlo simulation approach due to the explicit knowledge of the Jacobian function involved in normalizing the resulting MICA distribution. Nevertheless we have shown that even the computationally simpler optimization approach outlined in this chapter suffices to outperform classical ICA.

Tables 4.3 and 4.4 show the relative performance of MICA for contrast images and densely sampled textured regions, respectively. Given the original image I, the corresponding contrast image $J = C(I)$ was obtained as follows:

$$J(m,n) = \sqrt{\left(\sum_{i,j}^{N} w_{i,j}\right)^{-1} \sum_{i,j}^{N} \frac{w_{i,j}\left[I(m-i,n-j) - \mu(m,n)\right]^2}{\left[\mu(m,n)\right]^2}}$$

where N is chosen so that the contrast at each point in the image is computed in a 32x32 window [10] about (m, n), $\mu(m,n) = \sum_{i,j} w(i,j) I(m-i,n-j)$ is the local mean of image I around a 32x32 window about (m, n), and $\{w(i, j)\}$ is a set of raised cosine filter weights applied to the 32x32 window [70]. Contrast plays an important role in visual perception and is the basis for visual adaptation and other mechanisms employed by the human visual system in encoding low-level visual information [14] and for directing visual attention [125]. It is also a useful feature of image processing algorithms that seek to

emulate human performance [67]. Table 4.3 shows that MICA appears to outperform classic ICA when modeling the image patch statistics of contrast images.

Finally we also consider the situation where a 32 x 32 patch of a luminance texture image is densely sampled with M x M patches (M=3). The resulting samples (resulting in roughly N_{ptch} = 1024 samples/channel) are used to train the MICA model, as before, and subsequently compared with the ICA model. In Table 4.4 MICA again outperforms classic ICA in modeling the densely sampled image patch statistics.

We emphasize that all the results obtained above are for sub-optimal MICA distributions inasmuch as the parameter β was heuristically chosen, and the matrix B was not updated in subsequent iterations. Nevertheless consistent and statistically significant improvement relative to classic ICA are obtained when modeling image patch statistics, while at the same time revealing detailed quantitative information about the statistical interactions between the ICA components. We finally point out that even further improvements in the MICA model are likely possible by means of direct estimation of β parameter, incorporation of simulation phase of optimization, further refinement of the B matrix etc which in turn will be facilitated by the devising of faster and more efficient MICA parameter estimation algorithms.

These results demonstrate the considerable promise that multilinear modeling has in capturing the image patch statistics of natural images. Such models can find important applications in image processing and computational vision.

4.4 Discussion

In this chapter we have developed multilinear extension of ICA with application to the

modeling image patch statistics. A simple linear-quadratic non-linearity was shown to successfully account for dependences between the pseudo-ICA components, consequently approximating the true structure of the original joint probability distribution much better than possible with simple linear ICA. The quantitative information obtained about the statistical dependences between the pseudo-ICA components, which is naturally furnished by the MICA model, can potentially be used in a variety of applications such as non-stationarity measurement in natural images [76], texture synthesis, and modeling of simple cells in visual cortex.

Apart from such applications, there are open problems that emerged from this work of which we briefly mention a few:

(1) *Sparse Coding:* Consider a sparse coding problem involving the joint minimization of the MSE (i.e. mean-squared coding error with respect to $\{\phi_i\}_{i=1}^{d}$) and a sparsity term induced by $g(J)$. Is there an optimum basis set that is a solution to this problem?

(2) *Over-complete Models:* A first step is to address the problem of parameter estimation of a mixture of MICA models. This would have the added benefit of enabling the analysis of data from multi-modal probability distributions.

(3) *Non-sparse Multilinear Forms:* The basic methodology outlined here can be used to explore the original joint distribution with respect to projections on arbitrary basis; for example, the matrix B can be initialized with Gabor vectors.

Finally, there is considerable scope for improving the existing MICA model in terms of devising more efficient algorithms for parameter estimation, thus improving

parameterizations of the MICA model and thus for unleashing the full potential of this statistical modeling methodology.

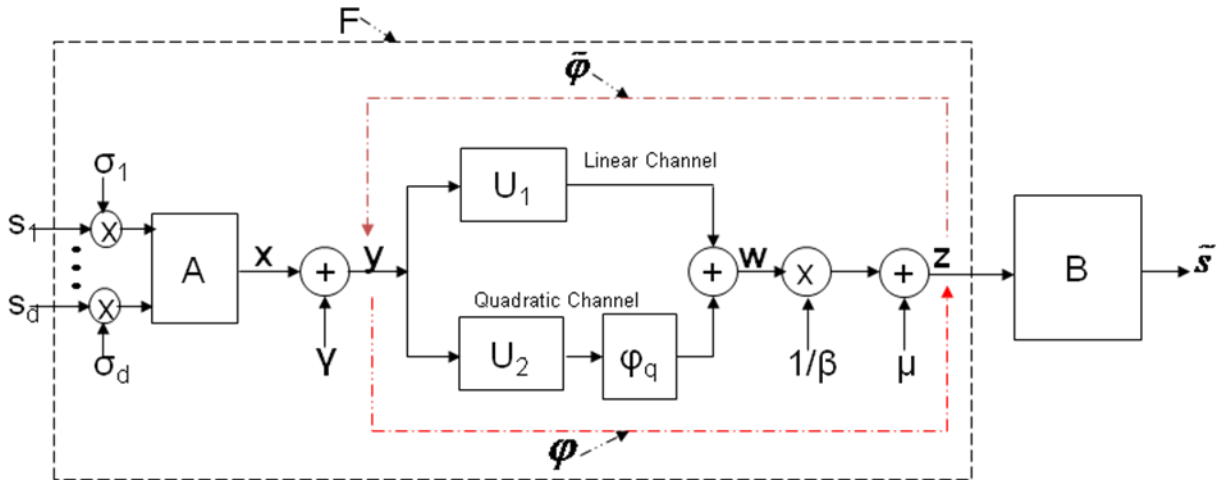
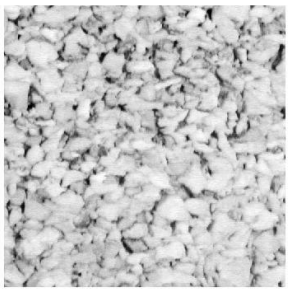


Fig. 4.1. Non-linear system model of the multilinear structure of source statistics derived from natural scene models.

(a)



(b)

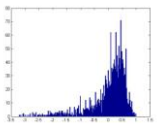
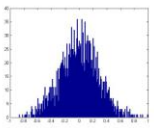
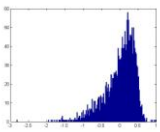
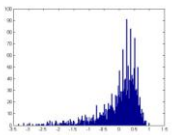
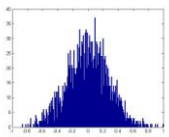
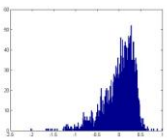
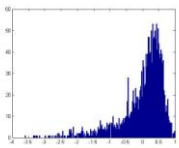
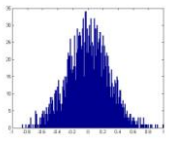
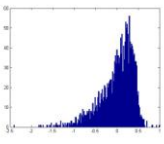
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.2 (*a*) Gravel (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The high-kurtosis heuristic $\beta_{high-kurt}$ was used.

(a)

(b)

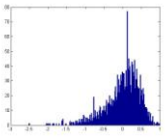
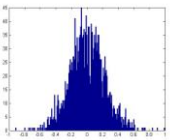
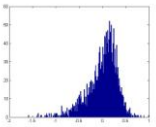
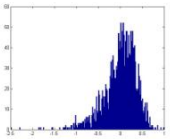
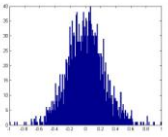
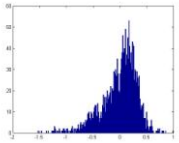
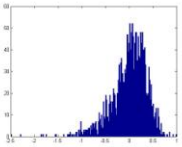
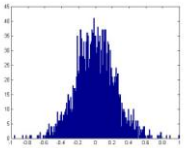
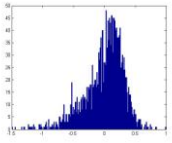
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.3. (*a*) Sand. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The high-kurtosis heuristic $\beta_{high-kurt}$ was used.

(a)

(b)

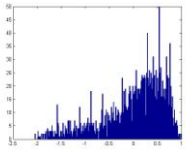
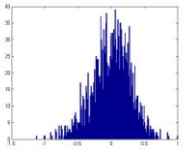
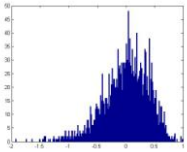
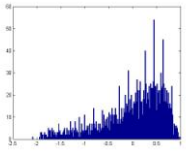
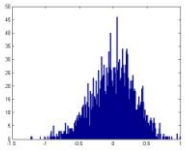
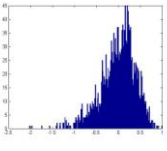
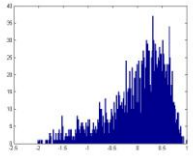
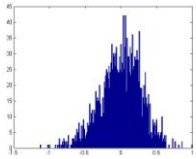
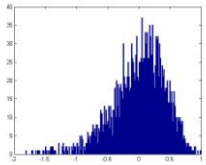
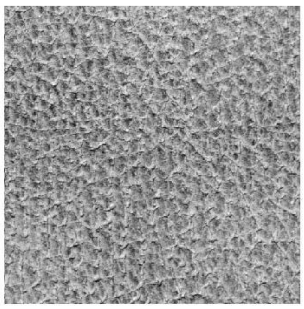
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.4. (*a*) Bark. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used.

(a)



(b)

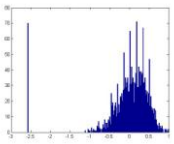
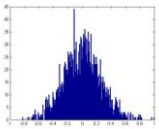
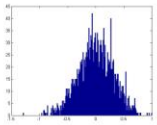
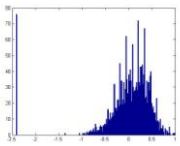
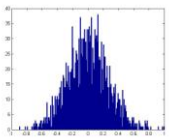
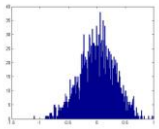
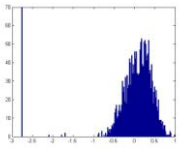
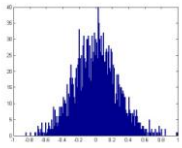
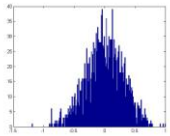
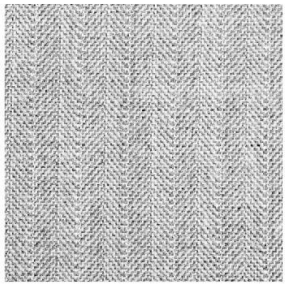
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.5. (*a*) Pigskin. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used.

(a)



(b)

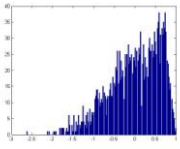
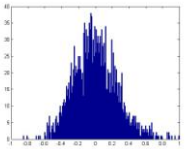
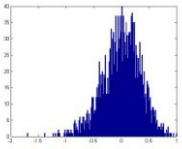
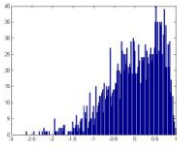
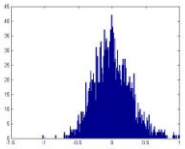
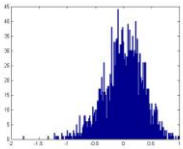
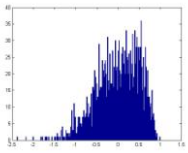
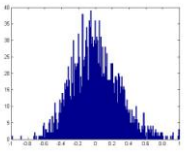
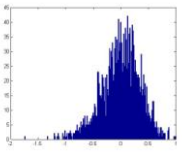
	Data	ICA	MICA	
Chan 1				
Chan 2				
Average of Channels				

Fig. 4.6. (*a*) Herringbone. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used.

(a)

(b)

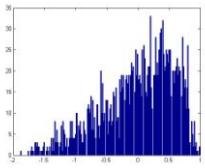
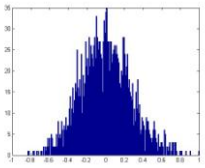
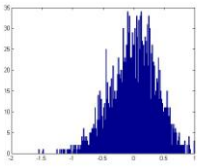
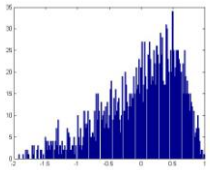
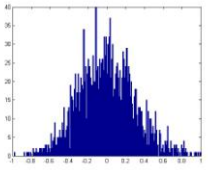
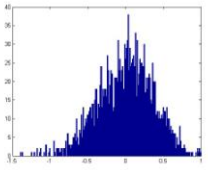
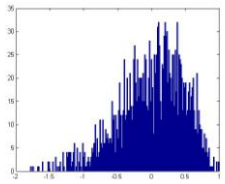
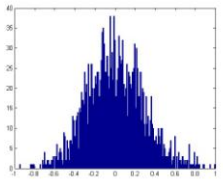
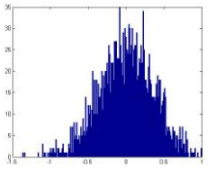
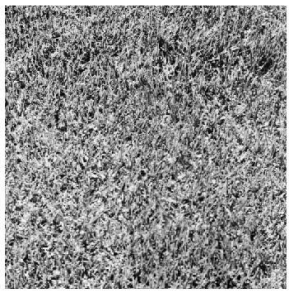
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.7. (*a*) Straw. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used.

(a)



(b)

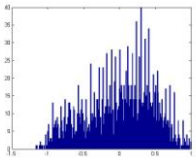
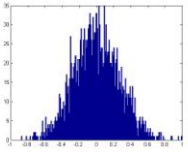
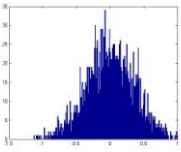
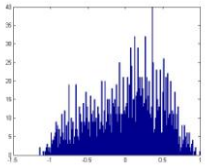
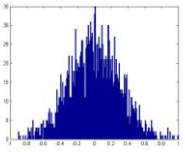
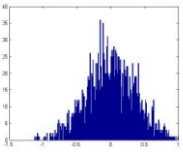
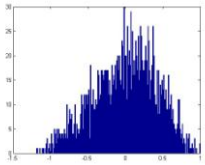
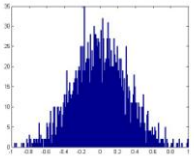
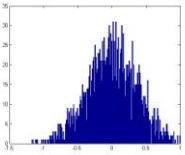
	Data	ICA	MICA
Chan 1			
Chan 2			
Average of Channels			

Fig. 4.8. (*a*) Grass. (*b*) Channel histograms of channels and their corresponding ICA and MICA distributions. The low-kurtosis heuristic $\beta_{low-kurt}$ was used.

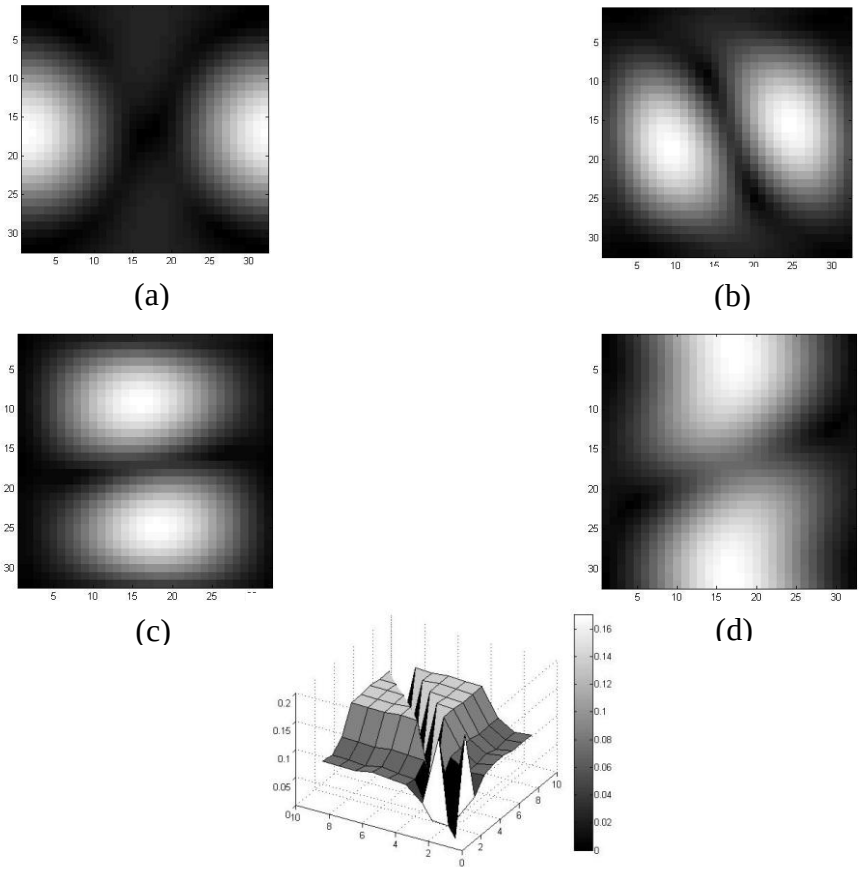


Fig. 4.9. (*a*)-(*d*) Examples of frequency responses of MICA filters corresponding to the Gravel texture; (*e*) magnitude |*G*| of the MICA interaction matrix for the Gravel texture. The larger the magnitude of $G_{i,j}$ the greater the statistical dependency between the corresponding MICA components.

High-level MICA Algorithm:

(1) Acquire data samples *x*
(2) Compute B-matrix in Fig. 1 (Initialization using Comon's algorithm [5], Gabor bases etc.; updation using Eq. (4.8))
(3) Compute the optimal MICA parameters using Eqs. (4.2)-(4.7)
(4) Repeat steps (4.2)-(4.3) as needed
(5) The optimal MICA parameters thus computed furnish the Multilinear model in Eq. 4.1

Fig. 4.10. High-level MICA algorithm

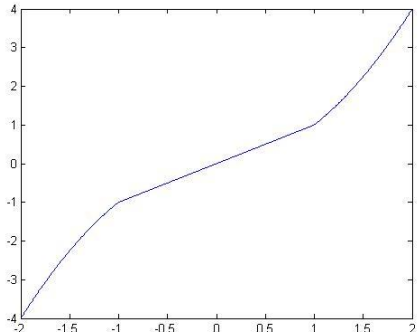


Fig. 4.11. Function φ: Linear in the unit interval and quadratic outside.

TABLE 4.1
%IMPROVEMENT IN KLD (W.R.T. ICA) DUE TO MICA MODEL

Texture	% θ_{KLD}^{MICA}
Gravel	46.8338
Sand	58.1021
Bark	41.5387
Pigskin	24.5900
Herringbone Weave	43.9410
Straw	39.9457
Grass	3.9546

Relative improvement with respect to classic ICA when using the *complete basis* MICA model is shown for the various textures

TABLE 4.2
%IMPROVEMENT IN KLD (W.R.T. ICA) DUE TO MICA MODEL

Texture	% θ_{KLD}^{MICA}
Gravel	24.4265
Sand	40.6984
Bark	31.5717
Pigskin	14.5383
Herringbone Weave	7.7534
Straw	19.7296
Grass	20.8215

Relative improvement with respect to classic ICA when using the *under-complete* MICA model.

TABLE 4.3
%IMPROVEMENT IN KLD (W.R.T. ICA) DUE TO MICA MODEL

Texture	% θ_{KLD}^{MICA}
Sand	29.5017
Herringbone Weave	48.4710
Straw	27.2106

Relative improvement with respect to classic ICA when using MICA for *contrast images*

TABLE 4.4
%IMPROVEMENT IN KLD (W.R.T. ICA) DUE TO MICA MODEL

Texture	% θ_{KLD}^{MICA}
Sand	29.5017
Herringbone Weave	48.4710
Straw	27.2106

Relative improvement with respect to classic ICA when using MICA for densely sampled texture regions

Chapter 5

Non-stationarity Measurement in Natural Images

5.1 Introduction

We introduce and develop the concept of non-stationarity indices which will prove useful in formulating optimal texture-based fixations in the next chapter. Our approach takes the view that since natural images are generally non-stationary, characterizing the non-stationary structure of images may yield useful insights into identifying regions of high information.

Non-stationarity is a ubiquitous feature of natural image data. Even simple spatial patterns corresponding to natural image textures such as grass, foliage etc, which are often modeled as spatially stationary processes for simplicity and tractability, generally contain varying degrees of spatial non-stationarity. The variations of non-stationarity across an image are often particularly significant in regions that occupy the transition between textures. Though the concept of non-stationarity is well founded and visually evident in natural images—which therefore enable us to make qualitative statements such as those given above—till now there has been no serious attempt at quantitatively measuring the extent of non-stationarities in natural images.

Most prior work in this area (and more generally in statistical image processing) has focused on modeling spatially non-stationary processes. The most common example is that of modeling 'roughly stationary' spatial processes called *textures*, for which a variety of different statistical modeling approaches have been proposed [96, 126-128] with varying degrees of success. More generally, it is often hypothesized that natural images

are comprised of non-linear combinations of (roughly) spatially stationary texture elements, where the non-linearities could be induced by occlusion, a simple partitioning into non-stationary textures, or other phenomena. With this hypothesis, a natural goal is to decompose the image into homogenous regions—each of which is 'roughly stationary.' Each such region can be described using *a priori* statistical models—for example by parameterizing the roughly stationary regions by MRF models [127]. This set of problems, which is generally referred to as texture-based image segmentation, has been an intense area of research for the last few decades [129]. Texture-based image segmentation is an extremely difficult problem for several reasons, of which the most central and important factor is the difficulty of ascertaining the non-linearities that account for the interaction of the texture elements. A step towards addressing this difficulty is to recognize that the effect of the non-linearities is to introduce non-stationarities in the image. Thus, the ability to accurately measure non-stationarities in natural images—which is the central topic addressed in this chapter—is the first step into gaining an insight into the interaction of the texture elements that comprise the image, which in turn may yield insights towards devising effective segmentation and identification algorithms for natural images.

The mathematical treatment of non-stationary processes has made important progress by the relaxation of the assumption of strict stationarity to enable the description of more general classes of spatial stochastic processes that are more interesting than simple models such as wide-sense stationarity (WSS) processes [153].

While the WSS assumption is overly simplistic for most image modeling applications; many real-world signals can be effectively modeled as approximately

stationary over localized spatial regions; for example the covariance matrix of the texture might vary significantly over the domain of the image yet change slowly among neighboring local image patches. Many different characterizations of the notion of local stationarity have been proposed in the literature [130-136]. Among these we single out the approach of Donoho *et. al.* [130-131] wherein the constructed class of locally stationary processes is diagonalized by local cosine basis functions, which is a natural generalization of stationary processes that are diagonalized complex exponentials. It turns out that efficient algorithms exist for the estimation of covariance matrices for this class of processes [130].

Another important direction of generalization is the construction of the so-called weakly harmonizable processes, wherein the covariance kernel is required to satisfy a certain generalized functional structure that reduces to the classical covariance functional (of WSS processes) under special circumstances. Weakly harmonizable processes were originally introduced by Bochner [137] as a generalization of earlier classes of stochastic processes constructed by Karhunen [138], Loeve [139] and Cramer [140]. Rao [141] showed that the weakly harmonizable class is the largest family of second order processes with continuous covariance for which Fourier analysis applies. In practical applications, recent work has shown promise in the modeling of linear time varying systems relating to multipath communication channels [142]. The application of the concept of harmonizable processes to the modeling of natural images remains, however, a relatively unexplored area.

A related and important body work involves the construction of explicit functional models of non-stationary processes to represent natural images. Important examples of

this include the fractional Brownian Motion (fBM) [143] and AM-FM models [136] of images. It is seldom the case, however, that natural images can be cast into any single class of non-stationary image model. AM-FM models apply to cases where the local frequency structure is assumed to vary slowly, which naturally excludes image regions that are locally very volatile and irregular which are better suited to fBM modeling (and vice versa).

Although important progress has been made into understanding the structure of non-stationary processes, previous approaches do not give a direct handle on ascertaining the degree of non-stationarity in natural images. There is a body of work called *change-point detection* [144] wherein the transitions between two stationary processes are detected. Most change-point methods, however utilize restricted classes of models that do not account for the actual statistics of natural scene data. Moreover none of these approaches quantify non-stationarity at every point in space—something we believe may prove to be an important ingredient for characterizing natural scenes.

In this chapter we directly address the problem of non-stationarity measurement in natural images (in Section 5.2). After a detailed theoretical treatment of this problem, we derive (in Section 5.2-C) a practical and computationally efficient non-stationarity index (the NANS Index) which we show to be effective for capturing the non-stationary behavior of images. We demonstrate the performance of the NANS Index for various natural, multi-texture and fingerprint images (in Section 5.3). We conclude with a discussion of important applications and directions of research that emerge from this work.

5.2 On Non-stationarity Measurement

A. Overview

We define an image region to be a set of contiguous pixels whose bounding contour (which we call a *window*) is a simple closed curve. A spatial random field is said to be stationary if, for an arbitrary window, the joint distribution of the random variables associated with the window remains invariant with respect to translation across spatial coordinates. The extent of non-stationarity across an image can thus be measured by, for example, the Kullback-Leibler Divergence (KLD) [116] between the relevant joint distributions corresponding to two different regions in the image. For a given realization of the random field (i.e. an image), however, the problem of determining the extent of non-stationarity across the image becomes much more difficult since these joint distributions, to which we have no direct access, must be estimated.

Two major issues arise when estimating joint distributions across a given natural image. The first is to precisely define the window structure employed to analyze image regions within the image—the size of resulting image regions analyzed determines the *image scale* of the statistical analysis. For a given image scale, the second problem is that of defining the probability space over which the statistical measurements (in particular non-stationarity measurements) are to be made. This involves choosing the random variables employed to characterize the statistics of the image region involved, and thereafter, quantifying the joint probability measure associated with them.

There are several approaches that have emerged in the literature for addressing the latter problem. Traditionally Markov Random Fields (MRF) have been used for modeling image statistics as determined by the neighborhood structure and the associated clique

potential (that determines the nature of the neighborhood interactions) [127]. MRF models, however, only crudely capture image statistics because of the simplistic nature of the clique potentials—such as first order derivative responses—that are typically employed. The FRAME model [126] takes this a step further by explicitly learning the image priors from the data. However this method still relies on a hand-picked set of pre-defined filters from which the image priors are built. Sparse coding techniques overcome this limitation by learning the image priors directly from the data in as parsimonious a manner as possible. For natural scenes, this is equivalent to performing an Independent Component Analysis (ICA) on the image patches due to the heavy tailed nature of the marginals involved [145].

In this chapter we employ a novel refinement of ICA called Multilinear ICA (MICA) [74-75]. The need for this new statistical tool arises since there are always dependences between the ICA components of natural image patches, which make the ICA decomposition only an approximation [119, 124]. The MICA decomposition refines the ICA model in a way that makes it possible to elegantly capture these dependencies. We further show that such a characterization of image patch statistics finds natural applications in quantifying non-stationarity across an image. In the next section we describe a MICA-based non-stationarity index in detail.

Regarding the issue of image scale, in this chapter we perform image analysis corresponding to a single image scale as defined by a specific center-surround window—described in more detail in the following sections—which we utilize to compute a dense non-stationarity map over a given image. This image scale was chosen to roughly correspond to the scale of analysis corresponding to the fovea in the HVS. One can in

principle perform similar analyses at multiple scales.

In the sequel we distinguish between theoretical and practical non-stationary measures. We define a theoretical non-stationarity index to be a functional of the joint probability distributions of sub-regions within an image region that is responsive to the non-stationarity of the image region. We demonstrate such a theoretical non-stationarity index (based on the MICA decomposition) in the next section.

The theoretical MICA-based non-stationary measure is, however, not computationally feasible. We therefore search for practical alternatives which, though not analytically equivalent in the general case, are equivalent in form under precise conditions. To this end, we present a practical non-stationarity measure (the NANS Index) in Section 5.2-B. The performance of the NANS Index for multi-texture, fingerprint and natural images is presented in Section 5.3.

B. Theoretical Non-stationarity Index

Our approach to non-stationarity measurement is based on the MICA decomposition of image patch statistics. In order to make this chapter as self-contained as possible, we briefly review MICA and show how it can be deployed to define a theoretical non-stationarity index.

We define the *M*x*M image patch statistics* of an *N*x*N* image region to be the joint distribution of the random variables corresponding to the pixels in the *M*x*M* patches that sample the larger *N*x*N* region of the image. Thus *M* determines the *scale* of the statistical analysis. Note that the scale is related to image scale inasmuch as we require that $M << N$ in order to be able to collect enough samples to make reliable statistical measurements. If

d is the effective dimensionality of the image patch space, then $d = M^2$ is the complete basis case which we consider first. Later we also consider $d < M^2$ corresponding to larger scales.

The pivotal idea is to characterize the image patch statistics for a given (M, d) using a multilinear expansion [74-75]:

$$P(X) = \frac{1}{Z} g(J) \prod_{i=1}^{d} P(s_i) \tag{5.1}$$

where $X = [X_1, \ldots, X_M]$, $J = [J_1, \ldots, J_M]$, $s_k = X * \phi_k$ where $\{\phi_k\}_{k=1}^{d=M^2}$ are the MICA filters, g: $J = [s_1, \ldots, s_d]^T$, and Z is a normalizing constant.

Of all possible multilinear expansions of the form (5.1) that could describe the source distribution, we are interested in the one that makes the representation of the source as sparse as possible, i.e., that minimizes the contribution of $g(J)$. In particular, we seek closed form approximations for such a $g(J)$. Such a multilinear form retains all the attractive properties of the ICA decomposition, while lumping the interactions of the filtered responses into the function $g(J)$. When $g(J)$ is separable with respect to the filter responses (or identity), (5.1) reduces to the classical ICA representation.

The *raison d'etre* of the MICA model is that for natural images patches, a perfect linear ICA decomposition of the image patch statistics can never be achieved: the linear model assumed in classic ICA is too restrictive and thus a suitable non-linear model must be found to account for the interactions between the pseudo-ICA components. We have shown in [74-75] that a simple linear-quadratic model can indeed account for the non-linear dependence between the observed sources in natural images. It emerges that the specific structure of the joint multilinear probability density has the form

$$p(s) = \frac{K}{|J(F)|} g(J) \prod_{k=1}^{d} p\left[\left(s_i - \mu_i\right)\beta_i\right]$$

where

$$p(s_i) = K_i \exp\left(-a_i \left\{\tilde{\varphi}\left[\beta_i\left(s_i - \mu_i\right)\right] - c_i\right\}^2\right),$$

K_i normalizes the probability density $p(z_i)$, $a_i = \frac{1}{2}\sum_{k=1}^{d} \frac{C_{k,i}^2}{\sigma_k^2}$, and

$$c_i = \gamma_i + \frac{\sum_{j \neq i} \sum_{k=1}^{d} \frac{C_{k,j} C_{k,i}}{\sigma_k^2} \gamma_j}{\sum_{k=1}^{d} \frac{C_{k,i}^2}{\sigma_k^2}}.$$

We model the interaction function $g(J) = \exp\left\{-\sum_{i \neq j} G_{i,j} \varphi\left[\beta_i (s_i - \mu_i)\right] \varphi\left[\beta_j (s_j - \mu_j)\right]\right\}$ where

$$K = \frac{\exp\left[-\sum_{k=1}^{d} \frac{(C_k^T \gamma)^2}{2\sigma_k^2}\right]}{(2\pi)^{d/2} \prod_{k=1}^{d} \sigma_k K_k} \text{ and } G_{i,j} = -\sum_{k=1}^{d} \frac{C_{k,i} C_{k,j}}{\sigma_k^2}.$$

We describe the non-linearity $\tilde{\varphi}(x)$ consisting of complementary linear and quadratic channels as follows. Given complementary step limiters $u_1(x) = u(x+1)-u(x-1)$ and $u_2(x) = 1-u_1(x)$, where $u(x)$ is the unit step function, then $\tilde{\varphi}(x) = x u_1(x) + \varphi_q^{-1}(x) u_2(x)$ with $\varphi_q^{-1}(\beta z) = \left[\sqrt{|\beta z_i|}\,\mathrm{sgn}\left(\beta z_i\right)\right]_{i=1}^{d}$.

The *MICA interaction matrix* $[G_{i,j}]$ captures interactions between the MICA components. In particular, when $[G_{i,j}]_{i \neq j} = 0$, the MICA components are independent. The parameter β determines the tradeoff between the linear and quadratic non-linear components of $\tilde{\varphi}$, μ adjusts the mean of the MICA distribution, and γ (along with β) determines the skew of the marginal distributions by asymmetrically assigning the linear

and quadratic components of $\tilde{\varphi}$ within the effective domain of the distribution.

The above MICA model was derived for the complete basis case. However, as shown in Chapter 4, the MICA model can be easily extended to the under-complete case ($d < M^2$). The difference is that the initial estimate of the matrix B consists of the d most prominent ICA components as determined by the initial ICA estimation algorithm [124], followed by optimization of the resulting d channels as above. In our experiments we have found that even one such iteration is enough to outperform ICA based methods without re-estimating B.

Given this characterization of image patch statistics, the following furnishes a theoretical non-stationarity index that measures the degree of non-stationarity present in a given image region.

Definition: Let T be an image region that is partitioned into two non-overlapping windows: a *center patch* and a *surround patch*. For example, the center patch might be a square patch in the middle of T, and the surround patch the rest of T. Let $\{\phi_i^c\}_{i=1}^d$ be the MICA filters computed from the center patch. Further, let I^c and I^s be random variables that correspond, respectively, to *d-dimensional* samples of the center and surround patches, $J_c = \left[I_1^c, ..., I_d^c \right]$, $J_s = [I_1^s, ..., I_d^s]$, where $I_k^c = < I^c, \phi_k^c >$ and $I_k^s = < I^s, \phi_k^s >$. The joint probability densities associated with J_c and J_s are given by μ^c and μ^s, respectively. Then the *theoretical non-stationarity index* is

$$\eta = \left| 1 - \frac{\sum_{m \neq n} E_{\mu^s} \left[< \tilde{\varphi}(I_m^s), \tilde{\varphi}(I_n^s) > \right] G_{m,n}^s + C_1^s + C_2^s}{\sum_{m \neq n} E_{\mu^c} \left[< \tilde{\varphi}(I_m^c), \tilde{\varphi}(I_n^c) > \right] G_{m,n}^c + C_1^c + C_2^c} \right| \tag{5.2}$$

where $[G^c_{m,n}]$ and $[G^s_{m,n}]$ are the MICA interaction matrices corresponding to J_c and J_s respectively, $C^c_1 = D\left[\mu^c(J_c) \| g_c(J_c)\prod_{k=1}^d p_c(I^c_k)\right]$ is the model mismatch measure, $C^c_2 = E\left[\ln\left(\prod_{k=1}^d \frac{p_c(I^c_k)}{\mu^c_m(I^c_k)}\right)\right]$ (with similar expressions for C^s_1 and C^s_2) and $D(p\|q) = \int p \ln\left(\frac{p}{q}\right)$ is the Kullback-Leibler divergence. Here μ^c_m and μ^s_m are the marginals associated with the ICA approximations of J_c and J_s respectively, and $q(J_c) = g_c(J_c)\prod_{k=1}^d p_c(I^c_k)$ $q(J_s) = g_s(J_s)\prod_{k=1}^d p_s(I^s_k)$ are the MICA distributions associated with J_c and J_s. ♣

Lemma 5.1: The theoretical non-stationarity index η given in (5.2) assumes a value zero if the image region T is stationary.

Proof: In the Appendix.

As shown in the appendix, (5.2) measures the relative change in mutual information between the center and surround patches with respect to the MICA filters of the center patch. Note that we are only guaranteed a sufficiency condition in Lemma 5.1.We remark that it is possible to devise alternative MICA-based theoretical non-stationarity measures that satisfy, in addition to Lemma 5.1, a necessary condition. One simple example is where we measure the relative change in mutual information of the surround patch with respect to MICA filters derived from the center *and* surround patches. A disadvantage of such an approach, however is that one, in principle, has to compute the MICA for both the center and surround patches ((5.2) only requires computing MICA from the center patch). Keeping this in mind, we find it a useful conceptual goal to attain (5.2).

Examining the non-stationarity index η in (5.2), we find that it consists of non-linear expectations of the projections of the *MxM* image patches onto various pseudo-independent directions corresponding to the MICA components of the center patch. It also requires computing C_1^c, C_2^c, C_1^s, C_2^s which account for normalization of the marginal and joint distributions. The computation of all these expressions is extremely difficult in practice, as it would require calculating the MICA interaction matrices of the center and surround patches at every image coordinate, making it infeasible to efficiently compute the theoretical non-stationarity index.

We note, however, that for the case where β is very small - so that $\tilde{\varphi}$ is linear—referred to herewith as the *low-β case*—the expectations in (5.2) reduce to simple correlations. Assuming that MICA closely approximates the true image statistics, we have $C_1^c \approx 0$ and $C_1^s \approx 0$. Furthermore, for the low-β case, C_2^c and C_2^s are constant, since the marginals reduce to Gaussian distributions. Thus for the low-β case, the theoretical non-stationary index (5.2) involves weighted summations of the covariance matrices corresponding to center and surround patches. In the sequel, the practical non-stationary measure we define (the NANS Index), in effect, places different weighting factors on the covariance matrices and thus is similar in form to the non-stationarity index in (5.2) for the low-β case. The ultimate justification for using the practical non-stationarity index (NANS) derives from its good empirical performance when applied to natural images.

C. The NANS Index: A Practical Non-stationarity Index

We have defined a MICA-based theoretical non-stationarity index that measures the degree of non-stationarity between adjacent image regions across an image. However, the

computational difficulties involved in computing this quantity prevent its practical deployment for assessing the non-stationary structure of natural images.

In order to simplify matters, we consider the low-β case wherein the numerator and denominator of expression (5.1) reduce to linear combinations of the correlations between the various ICA components. We thus seek heuristic and efficient ways of choosing the linear weighting factors of the correlation values in order to obtain a practical non-stationarity index that demonstrates good performance on natural images.

When analyzing the non-stationarity structure of images an important consideration is the image scale at which the analysis is performed, which is determined by the size of the analysis window. In this chapter, we have chosen to analyze the non-stationary structure of images at approximately the scale at which the foveola analyzes image regions assuming a viewing resolution of 1 arc minute per pixel[1]. Specifically, we deploy a *center-surround window* that is divided into two non-overlapping sub-regions (that partition the window): the *center* and *surround patches.* This center-surround window is used to compute the non-stationarity index at each point in the image.

Let $\{\phi_i\}_{i=1}^{d}$ be the ordinary ICA filters learned from the center patch when analyzing *M*x*M* image statistics for $M > \sqrt{d}$. Let I_c be the center patch and let I_s be the surround patch. Since the dimensionality of the subspace is assumed to be *d*, it follows that

$$E\left[\left\| I_C - \sum_{i=1}^{d} < I_C, \phi_i > \phi_i \right\|^2\right] \leq E\left[\left\| I_S - \sum_{i=1}^{d} < I_S, \phi_i > \phi_i \right\|^2\right] \tag{5.3}$$

[1]We envision non-stationarity analysis as a method of understanding human fixation patterns on natural scenes (as explored in detail in Chapter 6). This choice of resolution sets a baseline with which performance with respect to human vision can be compared. Another reason for this choice of image scale is that we have experimentally found that when computing ICA vectors (using Comon's algorithm [124]) for image regions much smaller than 32x32, reliable ICA vectors are not always obtained.

Analyzing both sides of the above expression:

$$LHS = E\left[\left\langle I_c - \sum_{i=1}^{d} < I_c, \phi_i > \phi_i,\ I_c - \sum_{i=1}^{d} < I_c, \phi_i > \phi_i \right\rangle\right]$$

$$= E\left[\|I_c\|^2\right] + \sum_{i=1}^{d}\sum_{j=1}^{d} E\left[< I_c, \phi_i >< I_c, \phi_j >\right] < \phi_i, \phi_j > - 2\sum_{i=1}^{d} E\left[|< I_c, \phi_i >|^2\right]$$

$$= E\left[\|I_c\|^2\right] + \|W \circ K_{center}\| - 2\text{trace}(K_{\text{center}})$$

$$= \left[\|W \circ K_{\text{center}}\| - \text{trace}(K_{center})\right] + \Delta E_{\text{center}}$$

where $\|\cdot\|$ is the L_1 matrix norm, $\circ$ denotes the Hadamard (point-wise) product, K_{center} is the covariance matrix corresponding to the center patch,

$W = [w_{i,j}]_{i,j=1}^{d} = [< \phi_i, \phi_j >]_{i,j=1}^{d}$, and $\Delta E_{\text{center}} = E\left[\|I_c\|^2\right] - \text{trace}(K_{\text{center}})$

A similar expression holds for the RHS:

$$RHS = \left[\|W \circ K_{\text{surround}}\| - \text{trace}(K_{\text{surround}})\right] + \Delta E_{\text{surround}}$$

where K_{surround} is the covariance matrix corresponding to the surround patch and

$$\Delta E_{\text{surround}} = E\left[\|I_s\|^2\right] - \text{trace}(K_{\text{surround}}) .$$

Thus we define the practical natural image non-stationarity index (NANS Index) as

$$\eta_{\mathrm{NANS}} = \left| 1 - \frac{E\left[\left\| I_S - \sum_{i=1}^{d} < I_S, \phi_i > \phi_i \right\|^2\right]}{E\left[\left\| I_C - \sum_{i=1}^{d} < I_C, \phi_i > \phi_i \right\|^2\right]} \right|$$

$$= \left| 1 - \frac{\|W \circ K_{\mathrm{surround}}\| - \mathrm{trace}(K_{\mathrm{surround}}) + \Delta E_{\mathrm{surround}}}{\|W \circ K_{\mathrm{center}}\| - \mathrm{trace}(K_{\mathrm{center}}) + \Delta E_{\mathrm{center}}} \right|. \qquad (5.4)$$

The numerator and denominator of the NANS Index η_{NANS} consist of linear combinations of correlations between the ICA components, as in (5.1) for the low-β case. We further observe from (5.3), that in order to compute η_{NANS}, it is not necessary to compute correlation matrices corresponding to the center and surround patches, but rather, just the linear coding distortions for the center and surround patches (with respect to the ICA components of the center patch).

D. Center-Surround and Boundary-Detecting NANS Indices

We now consider specific NANS Index measurement scenarios. Unless explicitly defined otherwise, the KLD between two probability densities p_1 and p_2 refers to max[$D(p_{1||}\, p_2)$, $D(p_2||p_1)$]–the max-KLD between p_1 and p_2.

The type of analysis window architecture we described in the previous section, where inner and outer patches are being used define the general class of what we will refer to as *center-surround NANS Index,* or *CS NANS Index.* Although for analytical convenience we presume that the analyzing image patch has a circular geometry, in practice we use a rectangular geometry for convenience of implementation. Of course, the center and

surround patches could be implemented to approximate concentric circles, if desired. In our simulations, we let the center-surround window be a 64x64 square region in which the center patch is the central 32x32 square sub-region within the patch. Shortly, we will define another NANS Index geometry suited to the detection of textural or stationarity boundaries.

It should be clear that the CS NANS Index, by construction, is suited for detecting *point* (*or quasi-singular*) *non-stationarities.* We define a quasi-singular non-stationarity as a locally occurring irregularity within a background texture which causes a measurable difference (with respect to the KLD) between the statistics of the center and surround patches. By implication, the scale at which the center-surround window analyses the image must be larger than the scale at which the quasi-singular non-stationarity occurs within the texture. Given a quasi-singular non-stationarity, the response of the CS NANS Index will be spread in the vicinity of the non-stationarity by an amount determined by the image scale at which the non-stationary image is being analyzed, i.e. the smaller the image scale, the smaller the spread.

A different class of non-stationarities that occur in natural images are *boundary-type non-stationarities* which occur between regions that are approximately stationary, but that display different statistical characteristics. The CS NANS Index does not peak at such boundary non-stationarities, and indeed, the response falls to near zero at a distance about equal to the radius of the center patch of the CS NANS Index. However, the index does peak at a distance approximately equal to the difference between the inner and outer patch radii. In principle, these observations could be used to adapt the CS NANS Index

for non-stationary boundary detection. However, we propose to take a more direct approach by defining a modified NANS Index.

The second type of NANS Index is defined as a family of center-surround windows $\{W_\theta\}_{0\leq\theta\leq\pi/2}$ parameterized by a rotation angle, where W_0 is an elongated window of dimensions (for example) $2N$xN. The window is defined as having vertical dimension that is twice the horizontal dimension, since it will be divided into an *upper patch* and a *lower patch* which will be used to compute patch statistics. In our simulations we will use on overall 64x32 window divided into 32x32 upper and lower patches. Statistical comparisons are then made between the upper and lower patches instead of between center and surround patches. The other members of the class $\{W_\theta\}_{0\leq\theta\leq\pi/2}$ are defined by rotating the $2N$xN window about its center by angle θ. In what follows we will only require the angles $\theta = 0$ and $\theta = \pi/2$, which conveniently avoids the need for interpolation following rotation.

The modified non-stationarity measurement method is then defined by making the same comparisons between the upper and lower (or left and right, when $\theta = \pi/2$) patches as were made in the CS NANS Index between center and surround patches. We refer to this modified index as the Boundary Detecting or *BD NANS Index*. Shortly we will develop a method for measuring boundary non-stationarities using two windows in the BD NANS Index.

First we shall qualitatively examine the responses of BD NANS Indices to quasi-singular non-stationarity. For simplicity take $\theta = 0$. The response of a BD NANS Index to a point non-stationarity will be small if it falls at the boundary between the upper and lower patches. The response would then increase quickly above and below the non-

stationarity. Thus, the response is oriented, with peaks away from the singularity. This makes the BD NANS Index less desirable for analyzing or detecting such non-stationarities as compared to the CS NANS Index.

Conversely, given an ideal (straight) boundary non-stationarity with orientation θ, we expect the BD NANS Index to yield a locally maximum response maximum response when an oriented window W_θ is used such that $\theta = \phi$, and when the window is centered on the non-stationary boundary. The response should fall towards zero as the window is moved away (perpendicular distance) from the boundary. However, this does not lead to an easy method for matching the analysis windows to local non-stationarity boundaries. The following arguments detail a way to accomplish this.

Given an ideal (straight) non-stationarity boundary oriented at angle θ separating two stationary regions T_1 and T_2 described by probability distributions p_1 and p_2 such that $D(p_1 \parallel p_2) \neq 0$, exact alignment of a BD NANS Index with the boundary constitutes a locally maximum response. A small rotation and displacement of the window relative to the boundary by angle ϕ and amount λ - ½ (depicted in Fig. 5.1) respectively, will yield a monotonic reduction in the BD NANS Index response by an amount approximately equal to

$$\Delta\lambda(\Delta\lambda + 1)\frac{4\phi}{\pi}(D_1 + D_2)$$

where $D_1 = D(p_1 \parallel p_2)$ and $D_2 = D(p_2 \parallel p_1)$.

To see that this is so, consider the generic situation of an idealized BD NANS Index deployed with a unit diameter circle window W_0. Further suppose that the windows lays across a straight line boundary non-stationarity oriented at an angle $\phi > 0$ relative to the

window, as depicted in Fig. 5.1. Suppose also that the window is laterally shifted by a distance ½-λ, where $\lambda \leq 1/2$ relative to the boundary (Fig. 5.1).

It follows that the probability distributions of the upper and lower patches (q_u and q_l) are approximately given by

$$q_u \approx \left(1-\frac{4\phi\lambda}{\pi}\lambda^2\right)p_1 + \frac{4\phi\lambda}{\pi}\lambda^2 p_2$$

$$q_l \approx \frac{4\phi}{\pi}(1-\lambda)^2 p_1 + \left[1-\frac{4\phi}{\pi}(1-\lambda)^2\right]p_2$$

Then it follows that

$$\Rightarrow D(q_u \,\|\, q_l) = \int q_u \ln\left(\frac{q_u}{q_l}\right) \approx \int q_u \ln\left(\frac{p_1}{p_2}\right)$$

The above approximations become exact as $\phi \to 0$. Then for small angles ϕ

$$H_\lambda^1 \equiv D(q_u \,\|\, q_l) = \left(1-\frac{4\phi}{\pi}\lambda^2\right)D(p_1 \,\|\, p_2) - \frac{4\phi}{\pi}\lambda^2 D(p_2 \,\|\, p_1)$$

Similarly,

$$H_\lambda^2 \equiv D(q_l \,\|\, q_u) = -\frac{4\phi}{\pi}(1-\lambda)^2 D(p_1 \,\|\, p_2) + \left[1-\frac{4\phi}{\pi}(1-\lambda)^2\right]D(p_2 \,\|\, p_1)$$

By construction, λ = ½ corresponds to the case where the center-surround window is centered exactly on the boundary non-stationarity, wherein:

$$H_{1/2}^1 = \left(1-\frac{\phi}{\pi}\right)D(p_1 \,\|\, p_2) - \frac{\phi}{\pi}D(p_2 \,\|\, p_1),$$

$$H_{1/2}^2 = -\frac{\phi}{\pi}D(p_1 \,\|\, p_2) + \left(1-\frac{\phi}{\pi}\right)D(p_2 \,\|\, p_1)$$

Let $D_1 = D(p_1 \,\|\, p_2)$ and $D_2 = D(p_2 \,\|\, p_1)$. Then, two distinct and exhaustive cases arise:

Case 1: (D_1-D_2<0) Here

$$H_{\text{ctr}} = \max(H_{1/2}^1, H_{1/2}^2) = D_2 - \frac{\phi}{\pi}(D_1 + D_2)$$

$$H_\lambda = \max(H_\lambda^1, H_\lambda^2) = D_2 - \frac{\phi}{\pi}(1-\lambda)^2(D_1 + D_2)$$

Perfect alignment with a boundary non-stationarity constitutes a local maxima, since $H_{\text{ctr}} < D_2$ and $H_\lambda < D_2$. But also $H_{\text{ctr}} > H_\lambda$ since

$$H_{\text{ctr}} - H_\lambda = \frac{\phi}{\pi}\left[4(1-\lambda)^2 - 1\right](D_1 + D_2) > 0 \,.$$

Case 2: (D_1-$D_2 \geq 0$) Here

$$H_{\text{ctr}} = \max(H_{1/2}^1, H_{1/2}^2) = D_1 - \frac{\phi}{\pi}(D_1 + D_2)$$

$$H_\lambda = \max(H_\lambda^1, H_\lambda^2) = D_1 - \frac{\phi}{\pi}\lambda^2(D_1 + D_2)$$

Again, perfect alignment with the boundary non-stationarity constitutes a local maxima since $H_{\text{ctr}} < D_1$ and $H_\lambda < D_2$. But it also follows that $H_{\text{ctr}} < H_\lambda$ since

$$H_{\text{ctr}} - H_\lambda = \frac{\phi}{\pi}\left(4\lambda^2 - 1\right)(D_1 + D_2) < 0$$

In both cases it follows that for a shift by amount $\Delta\lambda = \lambda - ½$, the change in the BD NANS Index due to an effective shift $\Delta\lambda$ in the radial direction is approximately

$$H_{\text{ctr}} - H_\lambda = \Delta\lambda(\Delta\lambda + 1)\frac{4\phi}{\pi}(D_1 + D_2) \,.$$

This establishes the approximation.

It is possible, given the preceding arguments, to employ a bank of BD NANS Indices in order to accurately detect texture boundaries with arbitrary orientations. However, it is computationally very intensive to do so. The following Proposition shows, however, that only two such BD NANS Indices —corresponding to $\theta = 0$ and $\theta = \pi/2$ — are sufficient to detect both quasi-singular and boundary non-stationarities.

Proposition 5.1: Let M_h and M_v be the BD NANS Index maps obtained by W_0 and $W_{\pi/2}$, respectively. Then $N = \max(M_h, M_v)$ is sufficient for detecting ideal boundary non-stationarities. ♣

Consider an ideal boundary non-stationarity oriented at angle $\theta < \pi/4$. Then minima occur at positions $\mathbf{x}_1 = (x_b - \sec\theta/2, y_b)$ and $\mathbf{x}_2 = (x_b + \sec\theta/2, y_b)$, where $\mathbf{x} = (x_b, y_b)$ is a location on the boundary. Between $\mathbf{x}_1$ and $\mathbf{x}_2$ the response will form a ridge, the magnitude and shape of which will depend on the probabilities p_1 and p_2. Once this ridge structure is identified, however, for each boundary location $\mathbf{x}$ by identifying $\mathbf{x}_1$ and $\mathbf{x}_2$, determining $\mathbf{x}$ along with the exact orientation angle θ directly follows.

When $\theta \geq \pi/4$, a similar argument applies except that the minima occur at locations $\mathbf{x}_1 = (x_b, y_b - \mathrm{cosec}\theta/2)$ and $\mathbf{x}_2 = (x_b, y_b + \mathrm{cosec}\theta/2)$. Again, since $\mathbf{x}_1$ and $\mathbf{x}_2$ can be measured, determining of the boundary location $\mathbf{x}$ and orientation angle θ directly follow.

In the next section we employ the NANS Indices derived above to study the non-stationary structure of various multi-texture, fingerprint and natural images.

One can in fact extend the above ICA-based NANS Indices (5.4) to arbitrary basis functions—such as PCA or Gabor bases—since a similar derivation will hold for such cases. But an advantage of an ICA-based approach is that one can get an approximate characterization of the probability density functions of the image patches by a product of the marginal distributions–and a more accurate characterization by means of the corresponding MICA decomposition [74-75]. This can lend a compact probabilistic and sparse characterization of the textural regions of the image.

Finally, although we have found it convenient to employ the KLD in the theoretical analysis of non-stationarities, it is difficult to actually compute the KLD in practice even given the MICA representations of image regions! Therefore, we have taken a more indirect route to measuring the degree of non-stationarity. In particular, the theoretical non-stationarity index involves computing the mutual information between the center and surround patches. Lemma 5.2 gives a relationship between the mutual information measurement and the KLD between the center and surround patches for a simple case.

Lemma 5.2: Let p and q be the probability densities associated with the center and surround patches in the CS NANS Index, respectively. Let $\{p_i\}_{i=1}^d$ and $\{q_i\}_{i=1}^d$ be the marginal distributions corresponding to the best ICA approximation of p and q, respectively. Further, let p be perfectly decomposable into its ICA components. Then

$$D\left(q \,\|\, \prod_i q_i\right) - D\left(p \,\|\, \prod_i p_i\right) = D(q \,\|\, p) + C \tag{5.5}$$

where

$$C = -\sum_i H(q_i) - \sum_i \frac{1}{\sigma_i^2} E_q\left[\left\{\tilde{\varphi}\left[\beta_i (x_i - \mu_i)\right] - c_i\right\}^2\right] - \sum_i \ln(K_i)$$ where H is the entropy functional, $\{\sigma_i, \beta_i, \mu_i, c_i, K_i\}$ are parameters associated with the MICA decomposition of p, and $\{x_i\}$ are filtered data with respect to the MICA filter bank associated with p. ♣

The constant C above is a linear combination of the various channel entropies and (possibly non-linear) variances. Lemma 5.2 thus gives a direct relationship between the mutual information measurements [LHS of (5.5)] and the KLD between the center and surround patches. The Lemma also applies to the BD NANS Index with small rewording.

More generally, the underlying intuition on which the NANS indices are based is that performing a MICA analysis on a non-stationary image region yields larger statistical dependencies than MICA analysis of the 'roughly' stationary sub-regions that comprise the image region.

5.3 Simulation Results

Given an NxN image region, the CS NANS Index is computed at every point within the region. To do so, at each point the center patch of the CS NANS Index is densely sampled with MxM windows. The resultant data vectors are then analyzed using Comon's algorithm [124] to obtain the ICA vectors characterizing the center patch. Next the coding errors corresponding to the center and surround patches with respect to these ICA filters are computed from which the CS NANS Index is computed using (5.4). A similar approach is used to compute the BD NANS Index, but substituting the relevant patches. Unless otherwise stated, all the BD NANS Index results are computed by evaluating the maximum of the BD NANS indices obtained by swapping the roles of the upper and lower patches at each point in the image. This is analogous to the max-KLD that we described in the previous section. In order to speed up computation, the NANS maps were evaluated on a sub-grid corresponding to a sub-sample factor of four along both rows and columns. The final NANS non-stationarity maps were obtained by performing an interpolation operation on this sub-grid.

Figures 5.2(*a*) and 5.3(*a*) show two different multi-texture images. The corresponding CS NANS Index maps (for $M = 5$, $d = 9$) are shown in Figs. 5.2(*b*) and 5.3(*b*), respectively. Higher values of the Index are observed near the boundaries, but with some

expected offset, as described earlier.

Also shown are the results of applying the BD NANS Index to the multi-textured images. Figures 5.2(*c*) and 5.3(*c*) show the result of applying the BD NANS Index with orientation $W_{\pi/2}$, while Figs. 5.2(*d*) and 5.3(*d*) show the result of applying the BD NANS Index with orientation W_0 to the same images. In each case we see the satisfying result of oriented BD NANS Indices responding strongly to boundary non-stationarities. As discussed earlier in Proposition 5.1, by combining orthogonally oriented NANS Index maps, complete boundary non-stationarity maps can be obtained, as shown in Figs. 5.2(*e*) and 5.3(*e*).

Figures 5.4(*a*)-5.5(*a*) show different fingerprint images on which we also apply the NANS Indices, but with some preprocessing. Fingerprints are of interest since they contain patterns whose local properties, such as orientation, change significantly over space. In some places the change is more rapid than in others, suggesting varying degrees of non-stationarity. Figures 5.4(*b*) and 5.5(*b*) depict the CS NANS Index maps from the fingerprint images. Figures 5.4(*c*) and 5.5(*c*) depict the result of applying the BD NANS Index with orientation $W_{\pi/2}$, and Figs. 5.4(*d*) and 5.5(*d*) with orientation W_0. Finally, Figs. 5.4(*e*) and 5.5(*e*) show the combined results using Proposition 5.1. Further, for visualization purposes, the non-stationary values were clipped to a value of five for all cases. We observe in Figures 5.4(b), 5.5(b) that the non-stationary index values tend to be high near the whorls of a fingerprint image since these are the places where there are more rapid changes in the spatial orientations of the locally sinusoidal-like line patterns. As we would expect, the non-stationary values tend to be high near the boundaries of the fingerprint structure. Furthermore we observe that the NANS index is also sensitive to

non-stationarities caused by isolated line and point-like structures in the image.

Finally we show the non-stationary structure of some natural images in Figs. 5.6-5.8. In each case the natural images were taken from the van Hateren database [13]. Each image contains foliage along with some man-made structures. In each case the processing by the NANS Indices is in the same sequence as the preceding examples. We observe that the CS NANS Index maps yield larger values where there is a transition between roughly stationary regions—for example in cases where man-made structures intrude upon the fauna, the Index is sensitive to non-stationarities created by sudden structures induced by point, edge and occlusion sources. On the other hand the Index map values are uniformly much smaller over the foliage regions, which despite changes in orientation, are qualitatively stationary at the given scale of statistical analysis. Similarly, the BD NANS Indices exhibit the directional behavior that is expected using both single windows and combined outputs. High responses are computed near sudden, sustained changes in stationarity, as can be seen between boundaries of man-made objects.

It should be noted that variations in the NANS Index values across an image are not generally reflective of sustained intensity changes, i.e. the NANS Indices are not "edge detectors." Rather they can be viewed as higher-level analogues wherein sustained statistical changes across the image are gauged with respect to the scale of image analysis.

5.4 Discussion

In this chapter we developed a computational theory of non-stationarity measurement in natural images. Though various theoretical treatments of non-stationary processes exist,

we believe that the fundamental way to make progress in *measuring* the degree of non-stationarity in an image is to employ appropriate NSS models to develop non-stationarity indices. In the pursuit of these goals we demonstrated a theoretical non-stationarity index based on our recently developed MICA model [74-75]. Thereafter we derived a more efficient non-stationarity index called the NANS index which, as we showed, has the same form as the theoretical non-stationarity index under certain conditions. The simulation results demonstrate that the NANS Index is sensitive to non-stationarities induced by various types of image structures, including those corresponding to point, edge and occlusion sources.

Though the NANS index (5.4) was derived for the ICA basis vectors corresponding to the image patch statistics, it can be extended to arbitrary basis functions—for example PCA or Gabor bases—since a similar derivation holds for such cases. Thus we can obtain a family of NANS non-stationary index functionals corresponding to different basis functions. In Chapter 6 we deploy a Gabor-based NANS Index and use it as a visual cue to compute visual fixations in natural images. In particular, we show that coupling the Gabor-based NANS Index with contrast statistics of natural images (as developed in Chapter 3) results in robust fixations patterns that correlate strongly with human visual fixations on grayscale natural images. This suggests that the human eye is attracted to the non-stationary structure of natural images when guiding visual fixations. Exactly how the NANS Indices can be used to compute features for higher level visual processing is a subject of future research. Apart from this broad research direction, many immediate problems arise – for example, whether one can devise efficient algorithms to directly compute the MICA interaction matrix. Doing so would enable direct computation of the

theoretical non-stationarity index, which would serve as a benchmark for assessing NANS Indices using arbitrary basis functions.

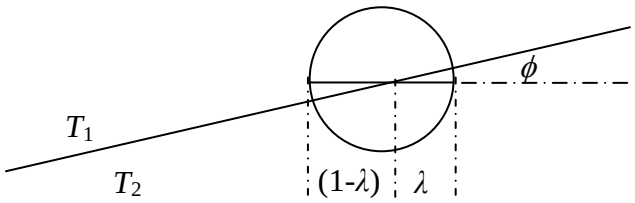
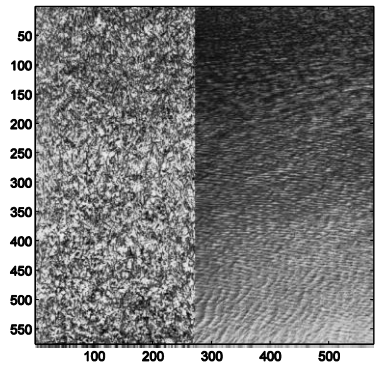
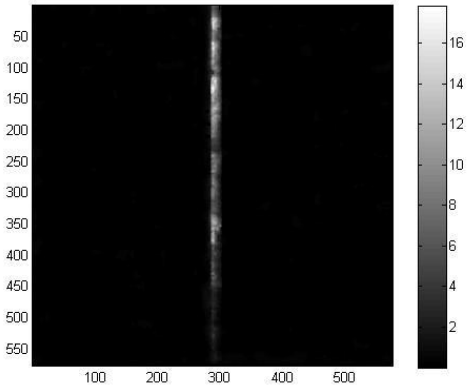


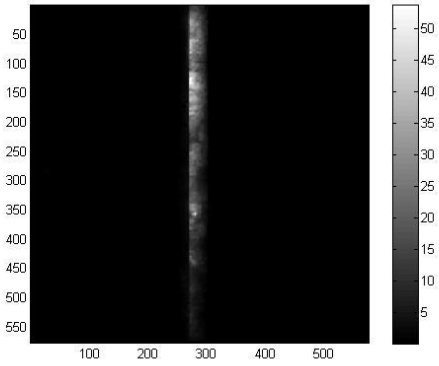
Fig. 5.1. Depiction of (unit diameter) BD NANS window with semi-circular upper and lower halves with rotation and offset from an ideal straight-line boundary non-stationarity.



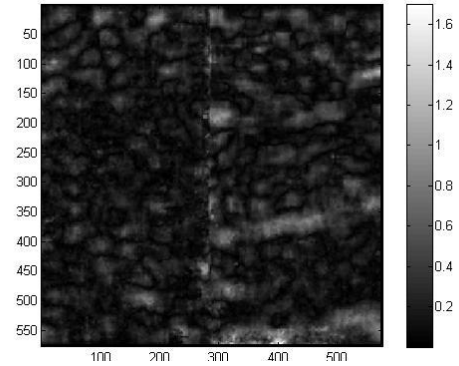
(*a*)



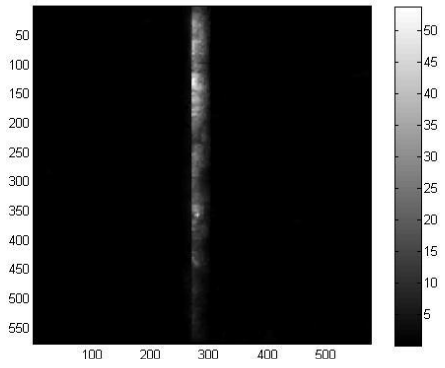
(*b*)



(*c*)

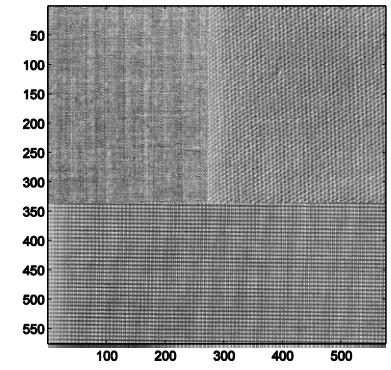


(*d*)

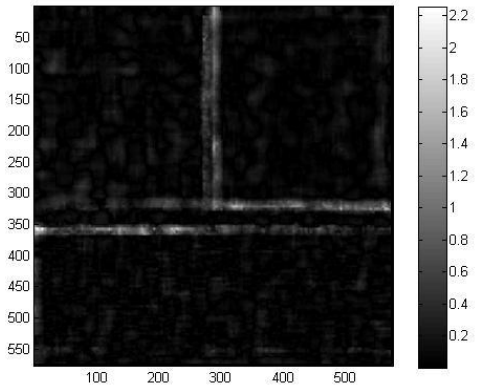


(*e*)

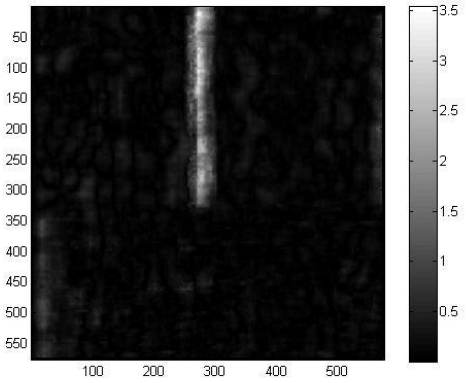
Figure 5.2. NANS processing of a multi-texture image. (*a*) multi-texture image; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1.



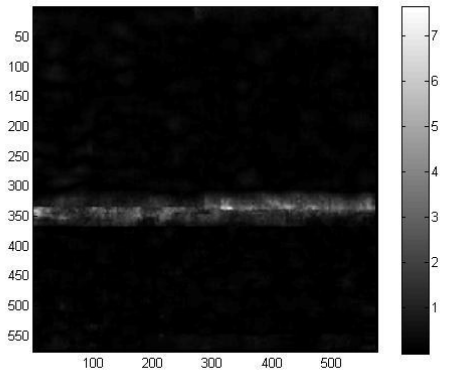
(*a*)



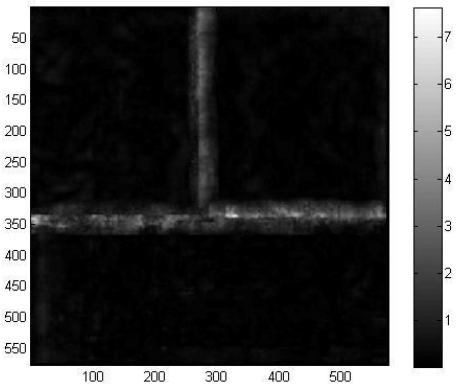
(*b*)



(*c*)

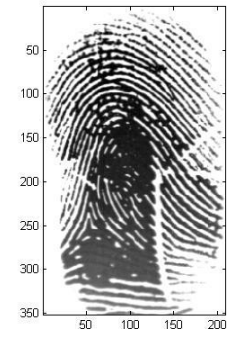


(*d*)

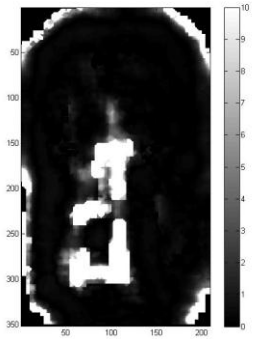


(*e*)

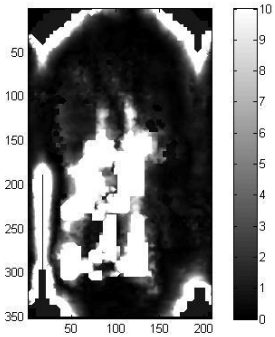
Figure 5.3. NANS processing of a multi-texture image. (*a*) multi-texture image; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1.



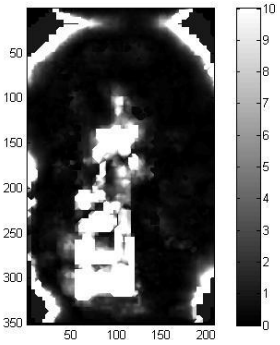
(*a*)



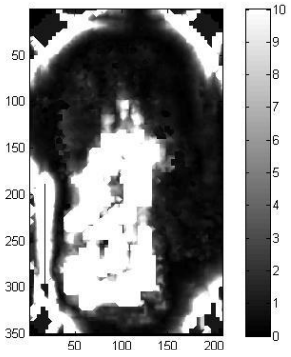
(*b*)



(*c*)



(*d*)



(*e*)

Figure 5.4. NANS processing of a fingerprint image. (*a*) fingerprint; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1.

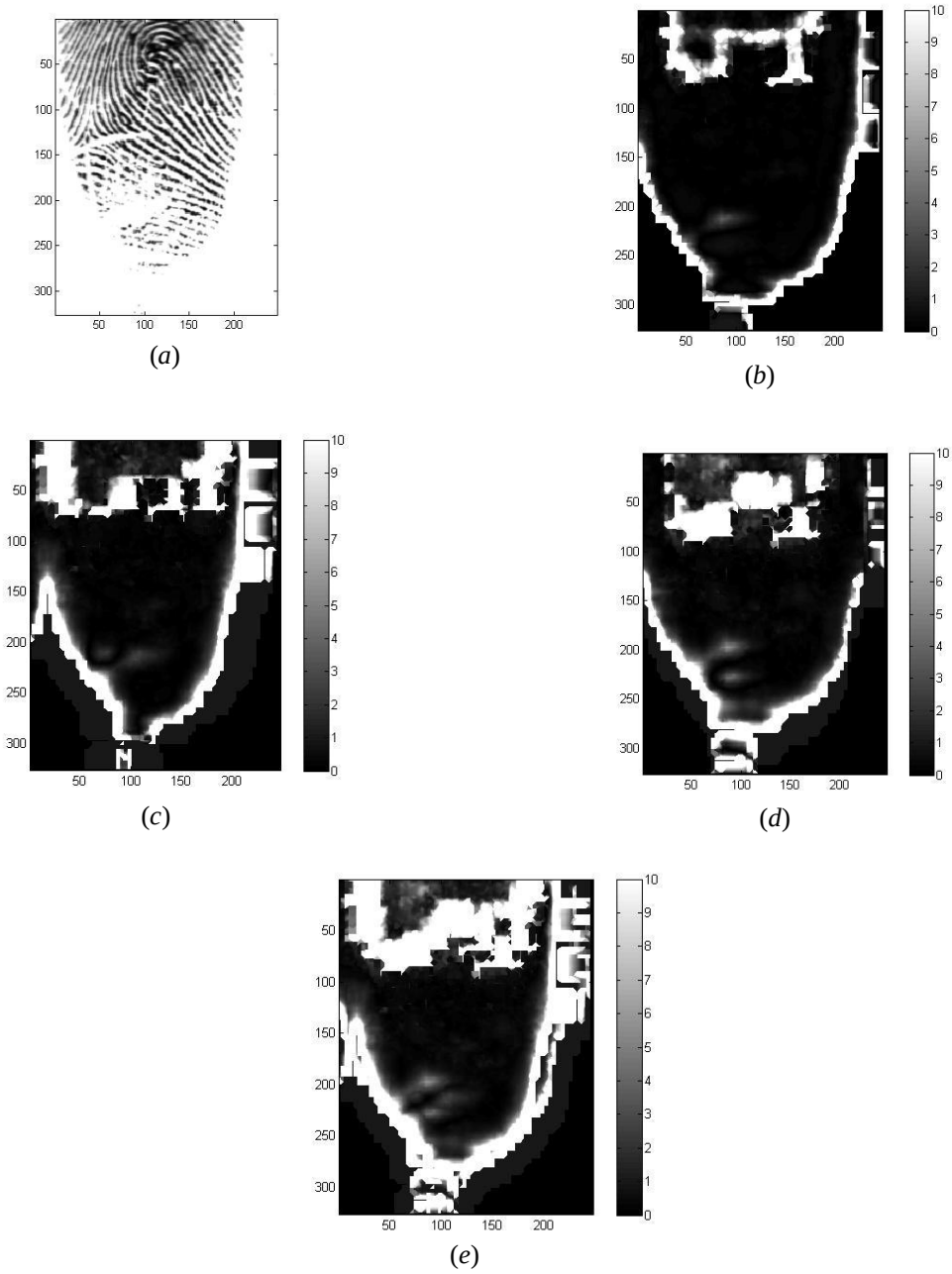
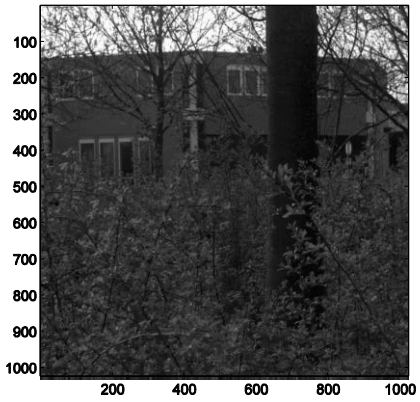
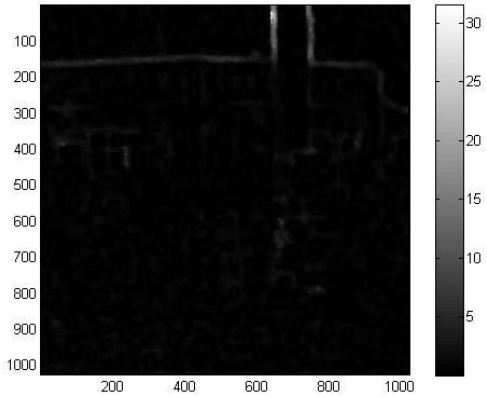


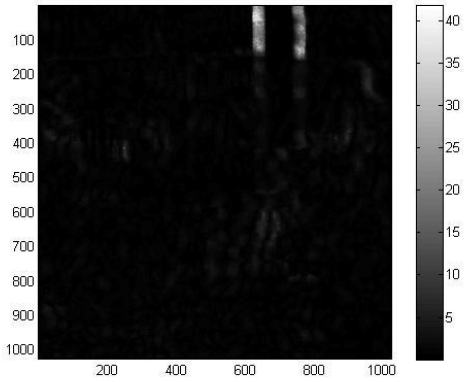
Figure 5.5. NANS processing of a fingerprint image. (*a*) fingerprint; (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1.



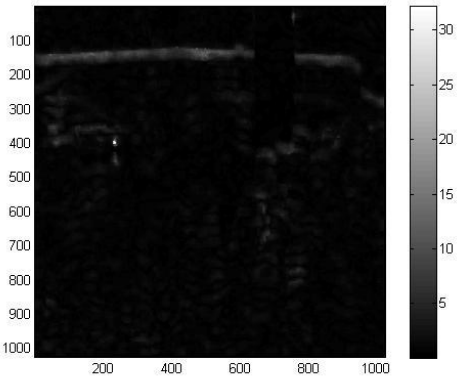
(*a*)



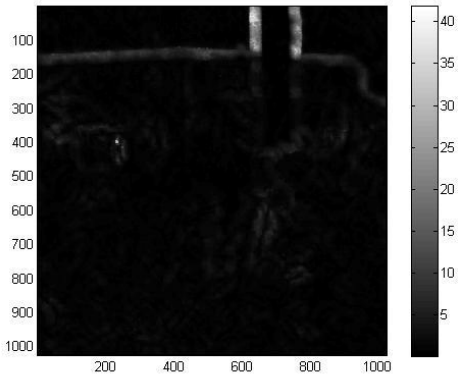
(*b*)



(*c*)

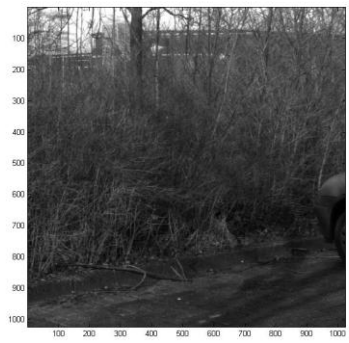


(*d*)

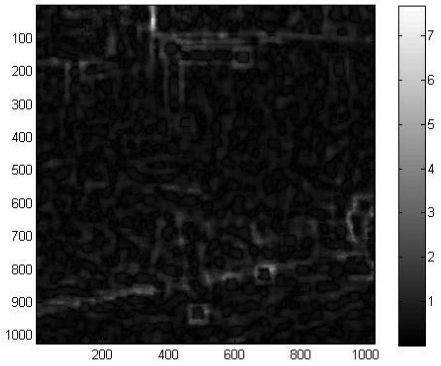


(*e*)

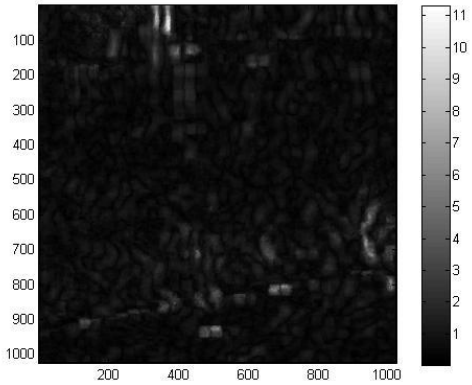
Figure 5.6. NANS processing of a natural image. (*a*) van Hateren image #1122 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown.



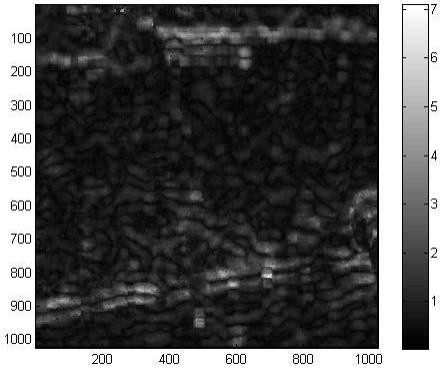
(*a*)



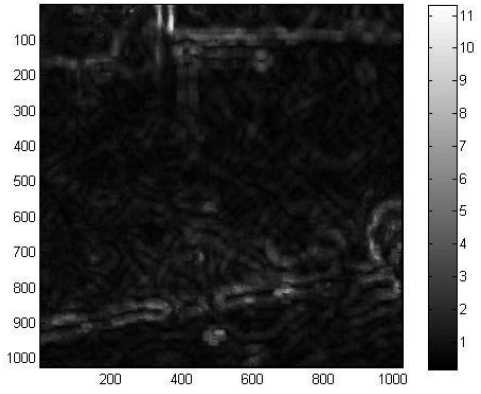
(*b*)



(*c*)



(*d*)



(*e*)

Figure 5.7. NANS processing of a natural image. (*a*) van Hateren image #8 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown.

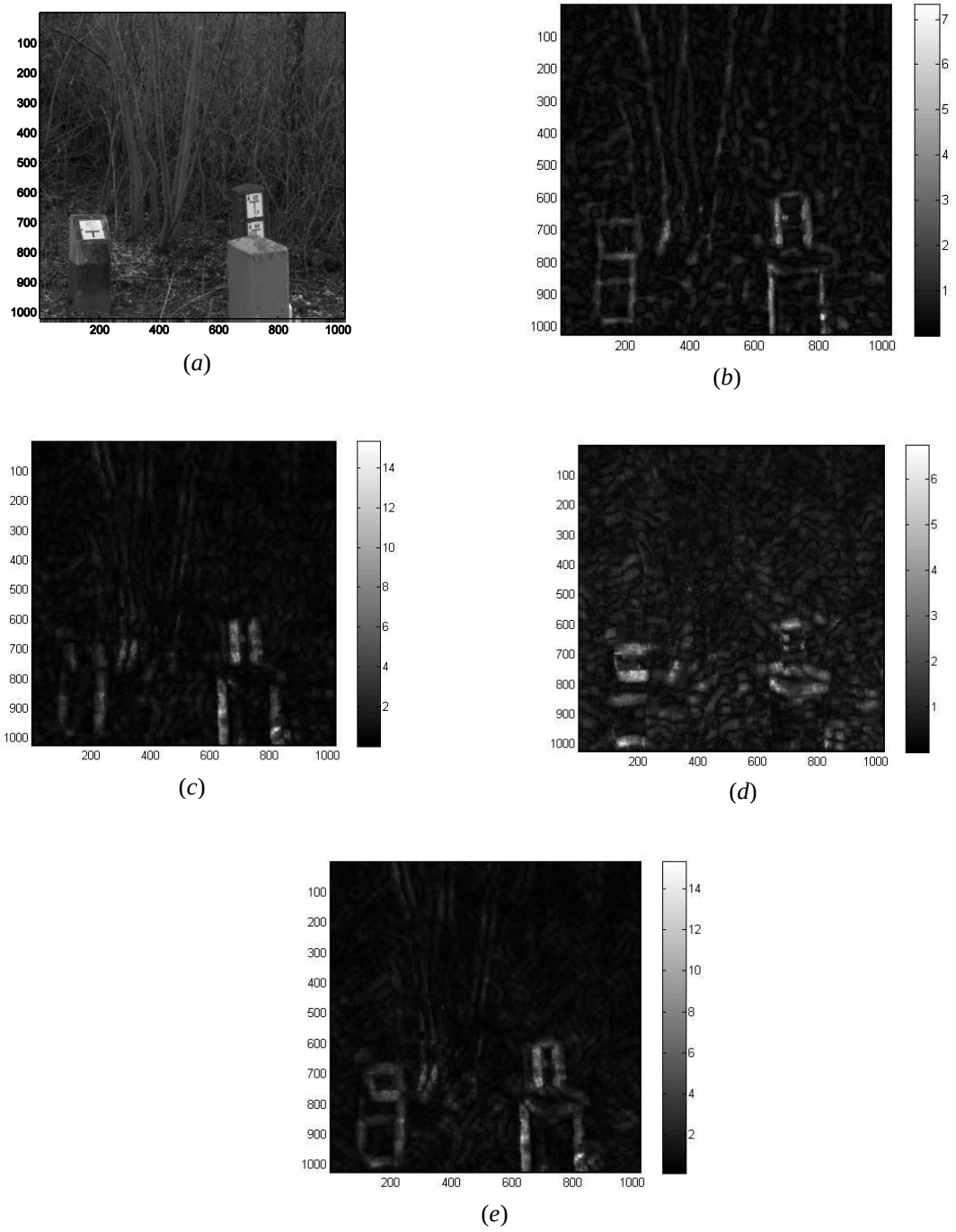


Figure 5.8. NANS processing of a natural image. (*a*) van Hateren image #93 (*b*) CS NANS Index map; (*c*) BD NANS Index map using $W_{\pi/2}$; (*d*) BD NANS Index map using W_0; (*e*) combined BD NANS Index map using Proposition 5.1. For display purposes the square-root of the NANS maps are shown.

Chapter 6

Texture-Contrast Based Fixation Selection in Natural Images

6.1 Introduction

The bewildering complexity of natural scenes is rivaled only by the amazing ability of the Human Visual System (HVS) to comprehend it. Comprehension, from an operational point of view, entails, in part, the systematic analysis and integration of different types of visual information at various levels of processing performed by the HVS—from low-level vision (corresponding to the 'front-end' of the HVS) to high-level visual processing (i.e. the 'back-end' processing of HVS)—and, of course, the subsequent utilization of the resulting knowledge to yield intelligent behavior. From a image processing point of view, it seems very reasonable that understanding of the workings of this complex system should also involve understanding of the nature of the information that the HVS is 'designed to process' at various levels of abstraction from low- to high- level processing. This point of view of course makes the tacit assumption that the HVS is optimized in some way to process visual information.

Attneave [1] and Barlow [2] hypothesized back in the 1950's that information theory can provide a link between environmental statistics and the properties of neural responses, in that the retina and other stages of the early visual system have evolved to develop efficient codes (i.e. in the least number of bits) for the information processed at the respective stages (given biological constraints at each stage such as the available number of neurons etc). Verifying the hypothesis entails not only the discovery of rich Natural Scene Statistics (NSS) models but also establishing precise quantitative

relationships to neural coding procedures that purportedly optimize certain aspects of NSS. Doing so would precisely establish the nature of the duality between NSS and low-level HVS processes.

Given the scope and generality of this hypothesis, various modified and restricted versions of this 'efficient coding hypothesis' have been proposed and verified by researchers [3-7]. More recently, work in the above two-fold research program of developing powerful theoretical models for NSS coupled with investigations into their implications for information processing in the HVS [8-13] have greatly advanced.

In this chapter we, for the first time, explicitly propose and verify a Barlow-type hypothesis for fixation selection in natural images. Our general hypothesis is that low-level visual fixations performed by the HVS in natural scenes are driven by the goal of maximally extracting visual information from the scene. We do not verify this hypothesis in its full generality but rather for the specific cases of textural and contrast information. After a brief review of optimum contrast based fixations in Section 6.2 (developed extensively in Chapter 3), we proceed, in Section 6.3, to develop an optimum texture-based fixation strategy based on our computation theory of non-stationarity detection which we developed Chapter 5. These two strands of work give us visual fixation patterns that optimally extract, respectively, contrast and textural information from natural scenes. We propose a simple coupling of these two fixation schemes and evaluate the performance of the resultant algorithms, in Section 6.5, by means of comparison to randomized fixation strategies via actual human fixations performed on the images. We find that the fixation patterns thus obtained substantially outperform both randomized and GAFFE-based [67, 146] fixation strategies in terms of matching human fixation patterns.

One of the important factors that motivates eye movement and visual fixations is that the HVS is a foveated visual system: the sampling density is highest at the point of fixation and gradually decreases from there [14-16]. At any fixation, the image acquired by the HVS contains less detailed information in the periphery. To acquire peripheral information at high resolution, the eye makes rapid ballistic movements – saccades. Conversely, foveation dramatically reduces the amount of information processed at each fixation.

The study of eye movements and fixation selection in humans is complicated by the fact that it is influenced by both top-down (high-level/cognitive) factors and a variety of low-level image features (i.e. bottom-up factors). In fact, identification of the types of image information that are important for the HVS—i.e. what visual cues, in low-level vision, exhibit statistical regularity that can be effectively exploited by the HVS; or what visual cues, in high-level vision, lend to useful conceptualizations that are relevant to the tasks at hand—is one of the major challenges in uncovering the nature of visual fixations performed by the HVS.

Being able to identify locations that are likely to attract human visual fixations is an important goal from a computational perspective for two reasons: first, it is a natural way to select regions for specific and more intensive processing, such as selecting quantization parameters in video compression [81, 147]. Secondly, future robotic vision systems are likely to deploy motile cameras that will be able to exploit the significant efficiencies enabled by foveation-fixation-based processing [66]. The processes that govern human eye movements represent excellent and efficient systems to emulate, at least as a start.

Early studies on eye movements by Yarbus [37] revealed that visual fixations are influenced by high-level factors such as the nature of the specific task being performed. Top-down approaches are popular in computer vision because the problem can be intuitively formulated in terms of high-level features of the object such as shape, spatial relationships between objects, and so on. Wixson [38] proposed an 'indirect search' strategy using spatial relationships between targets and its surroundings to first identify an intermediate object (associated with the target) that is easier to find and then search in that region for the target. Since knowledge about cognitive mechanisms employed by the HVS during visual search is limited, top-down approaches usually incorporate *ad hoc* assumptions regarding what features will be of interest during fixation mechanisms.

On the other hand, there is ample evidence to demonstrate that a significant proportion of the fixations performed by the HVS is driven by low-level features. The sheer volume of human fixations performed—about 15,000 fixations/hour—makes it implausible that the HVS uses computationally intensive semantic scene information to make a majority of the fixations. One of the most emphatic illustrations of the limitations imposed by low-level vision on performance in visual search was demonstrated in [61], wherein the role of low level features in visual search was assessed by measuring variations in discrimination performance. The influence of high level factors in search was minimized by using constrained experimental conditions (for e.g., two-alternate forced choice experiments with spatially and temporally localized targets). They show that search results as measured by accuracy and speed of performance are indeed influenced by low level factors like loss of spatial frequency in the retina and contrast masking.

Bottom-up approaches to fixation selection assume that eye movements are probabilistically driven by low-level image structures. Proponents of this paradigm [62-64] propose computational models for human gaze prediction based on image processing algorithms that accentuate image features that are deemed relevant. A few reported studies on automatic visual search have examined fixation selection based on features such as contrast, edges, object similarity [65] or combinations of randomized saliency and proximity factors [66]. In an interesting study, Privitera & Stark [62] used a suite of algorithms such as detecting the presence of symmetry, center surround regions in images that resemble receptive field profiles, wavelets, contrast, and edges- per-unit-area to predict points of interest in an image. They compared these predictions with human eye fixations. The comparison of the predictions and human eye movements was accomplished by analyzing their spatial/structural binding (location similarity) and temporal/sequential binding (order of fixations). They report that around 50% of their computed fixations matched those of human observers. A recent and more comprehensive study of low-level fixations was conducted by Rajashekar *et. al.* [67] wherein point-of-gaze statistical analysis of visual fixations was coupled with foveated analysis of fixation points. Extending previous work [68] done in a non-foveated fixation framework, the authors in [67] demonstrate that points corresponding to human fixations exhibit higher values of contrast, bandpass contrast, luminance and bandpass luminance on average as compared to random fixations performed on natural scenes. Furthermore, they proposed a simple fixation selection strategy (named GAFFE) that linearly combines saliency maps corresponding to all these features. The resulting GAFFE-based fixations [67] outperform random-based fixation strategies in natural images with respect to the

correlation measure.

In this chapter, we choose to view the fixation/foveation process as an information gathering process, made efficient by efficient selection and processing of visual data. None of the previous work on fixation selection in natural images (top-down or bottom-up) approaches the problem from an information theoretic point of view. We make beginning steps in this direction wherein we demonstrate how elegant solutions can emerge that yield superior fixation performance for natural images—and in doing so, we also pave the way towards a unified information-theoretic understanding of low-level fixation processes in the HVS.

6.2 Contrast-based Fixations

In order to make this chapter as self-contained as possible, we briefly review optimal contrast-based fixation strategies. The reader is referred to Chapter 2 for a more detailed treatment of this topic.

A. Contrast Statistics of Natural Images

Certainly, contrast is among the most important low-level image features encoded by the HVS [14]. In fact, the HVS does not effectively operate on luminance images but rather on the corresponding contrast images [14]. Hence even at the outset it is highly plausible that acquiring contrast information from natural scenes is an important sub-goal for the HVS when making visual fixations. In the following we describe a method by which visual fixations are deployed in a manner such that the maximum amount of image contrast information is accessed by a sequence of fixations of given length.

The local (patch) root mean square (RMS) contrast of an image at pixel location j is:

$$C_{\mathrm{RMS}}(j) = \sqrt{\frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \frac{(I_i - \bar{I})^2}{(\bar{I} + I_{\mathrm{dark}})^2}}$$

where I_i is the intensity of the i^{th} pixel in the patch, w_i is a weighting function, and I_{dark} is the "dark light" parameter chosen to be 7 td (1 cd/m^2 assuming a 3 mm pupil), based on human photopic intensity discrimination data (this parameter has little effect on the measured contrasts since the mean luminances of the images were generally >> 1 cd/m^2).

To compute local contrast we defined a circular patch of N pixels about each pixel j, and we chose to use a raised cosine weighting function:

$$w_i = 0.5\left[\cos\left(\frac{\pi}{p}\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\right) + 1\right]$$

where p is the radius, (x_j, y_j) is the location of the center pixel of the patch, and (x_i, y_i) is the location of the i^{th} pixel of the patch. The same weighting function was used to compute the local mean luminance $\bar{I}$. In our simulations the diameter of the raised cosine window was taken to be 32 pixels (i.e., $N \approx 256\pi = 804$).

We studied the properties of the conditional contrast distributions $P\{c \mid c_b(\varepsilon_i)\}$ for natural images, where c and $c_b(\varepsilon_i)$ are the local contrasts of the unblurred original image and the blurred image at eccentricity ε (angular distance from fixation), respectively. These conditional distributions can be regarded as the posterior probability distributions of the unblurred contrast, given the observed blurred contrast.

Empirical measurements of the conditional contrast distributions were carried out on a database of over 300 calibrated natural images found in [13]. We found that the contrast statistics are well-characterized by a simple set of formulas. The conditional distributions

of contrast given the blurred observed contrast are accurately fit by skewed Gaussian distributions:

$$P(c_a \mid c_b) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l + \sigma_h}{2}\right)} \exp\left[\dfrac{-(c_a - u)^2}{2\sigma_l^2}\right] & \text{if} \quad c_a > u \\ \dfrac{1}{\sqrt{2\pi}\left(\dfrac{\sigma_l + \sigma_h}{2}\right)} \exp\left[\dfrac{-(c_a - u)^2}{2\sigma_h^2}\right] & \text{if} \quad c_a < u \end{cases} \tag{6.1}$$

where u is the mode and (σ_l^2, σ_h^2) are the variances of the two halves of the skewed-gaussian distribution relative to the mode (see Fig. 6.1). The parameters $(u, \sigma_l^2, \sigma_h^2)$ vary in a simple fashion with the blurred contrast c_b:

$$u(c_b, \varepsilon) = (k\varepsilon + 1)c_b \tag{6.2}$$

$$\left[\bar{\sigma}(c_b, \varepsilon)\right]^2 = (k\varepsilon c_b)^2 + \sigma_0^2 \tag{6.3}$$

where σ_0 is a small constant,

$$\bar{\sigma}(c, \varepsilon) = \frac{\sigma_l(c, \varepsilon) + \sigma_h(c, \varepsilon)}{2},$$

and $k = 0.1082$ is an empirically determined constant. See Chapter 2 for more empirical plots of the properties described above.

B. Optimum Contrast-Based Fixations

A consequence of the above characterization of the contrast statistics of natural images is that the entropy of the conditional distribution is given by a simple formula:

$$h\left\{P\left[c \mid c_b(\varepsilon_i)\right]\right\} = \frac{1}{2}\log_2\left[2\pi e\left[\bar{\sigma}(c, \varepsilon)\right]^2\right] \tag{6.4}$$

Equation (6.4) allows us to formulate a simple algorithm that optimally extracts contrast images from natural images. Specifically, a greedy optimization approach is

employed, where our aim is to find a sequence of fixation points r_1, r_2, r_3, …, where r_j = (x_j, y_j), such that the $(T+1)^{st}$ fixation maximally reduces the total contrast entropy:

$$r_{T+1} = \arg\max_{r} \left[H(r_0,...,r_T) - H(r_0,...,r_T,r) \right]$$

where

$$H(r_0,...,r_T) = \sum_{i=1}^{T} h_i(T)$$

$$h_i(T) = \frac{1}{2}\log_2\left(2\pi e \left\{ \left[k\varepsilon_i(T) c_i(T) \right]^2 + \sigma_0^2 \right\} \right)$$

is the total contrast entropy after the T^{th} fixation, summed over all n pixel locations in the image. In these expressions

$$\varepsilon_i(T) = \min_{t \le T} \varepsilon_{it}$$

is the smallest eccentricity obtained so far at location i, while $c_i(T)$ is the contrast observed at that eccentricity. To determine the best next fixation it is necessary to estimate what the contrast entropy will be at every pixel, for every possible next fixation:

$$h_i(T+1) = \frac{1}{2}\log_2\left(2\pi e \left\{ \left[k\varepsilon_i(T+1) \hat{c}_i(T+1) \right]^2 + \sigma_0^2 \right\} \right) \tag{6.5}$$

To evaluate (6.5), we first use (6.1) to compute the MAP estimate of the unblurred image contrast at pixel location i, then apply (6.1) again to estimate the blurred contrast $\hat{c}_i(T+1)$ after making a fixation to location r. The initial fixation point can be chosen arbitrarily, or according to some *a priori* distribution. In the simulations given later, the center pixel of the image was taken as the initial fixation point. Finally, we remark that the value of σ_0 is not important, as long as it is small and positive. Further details on the implementation of this algorithm are given in Section 6.5.

6.3 Texture-based Fixations

A. Overview

We seek a natural extension of the information-theoretic approach to modeling visual fixations to encompass image texture. Here we define texture as a 'roughly stationary' spatial process such that the degree of non-stationarity decreases with increasing scale of spatial analysis.

The structure of natural images is the result of complicated non-linear interactions between texture elements, where the non-linearities can be induced by occlusions, boundaries, spatial transients, and other phenomena. While contrast is a highly local image property, texture is a regional concept—requiring probabilistic descriptions on multi-dimensional spaces.

However, non-linearities usually induce non-stationarities in the image, which bear considerable information regarding the structure of the image. Therefore we may pose that visual fixations that seek to extract textural information from natural images should be driven by image non-stationarities. Indeed, non-stationarities usually occur at locations where maximum computational resources will be expended on expensive tasks such as segmentation and recognition.

Clearly, if there are no significant non-stationarities present in an image, then it may be considered as a single texture, and so, performing multiple fixations will yield little textural information beyond the parameters of the texture model. Moreover, since statistical texture models generally assume that texture samples are drawn from stationary processes, then recognizing stationary image regions is an important aspect of image information gathering. Texture-based segmentation is an obvious example of this [76].

Towards this end we have proposed a quantitative measure of non-stationarity called the *Natural Image Non-stationarity Index* (NANS Index), which we briefly describe in the next section, along with modifications towards developing a texture-based fixation-finding strategy. In the sequel we couple texture- and contrast-based fixation-finding strategies into a single algorithm that delivers robust fixation performance on natural images.

B. Gabor-based NANS Index

We define an image region to be a set of contiguous pixels whose bounding contour (which we call a *window*) is a simple closed curve. A spatial random field is *stationary* if, for an arbitrary window, the joint distribution of the random variables associated with the window remains invariant with respect to translation across spatial coordinates. The size of the window defines the *scale* of image analysis [76].

Consider the case wherein the non-stationarity analysis window consists of two non-overlapping regions that partition the window—one called the *center patch* and the other, the *surround patch*. This could consist of concentric circular and ring-shaped regions, for example, or square approximations to them. When such a geometry is used the non-stationarity measurement is called a center-surround or CS NANS Index, to distinguish it from indices computed using other geometries, such as side-by-side patches [76]. The center-surround window is then centered at every image coordinate (pixel) allowing computation of the CS NANS Index at every coordinate.

In order to measure non-stationarity at each coordinate, in principle probability distributions must be associated with the center and surround patches. Then, non-

stationarity can be measured by, a distance measure such as correlation [67, 146] or Kullback-Leibler divergence (KLD). In Chapter 5 we showed how the joint probability measures can be naturally defined via an multilinear ICA decomposition [74] of the center and surround patches. This construction ensures that the mutual information of the surround patch always exceeds that of the center patch given that the image patch is non-stationarity—and also, under special circumstances, that it is directly related to the KLD between the center and surround patches [76].

The central idea of CS NANS Index is to gauge the relative change of mutual information between center and surround patches [76]. Let p and q be probability densities associated with the center and surround patches respectively; and let $\{p_i\}_{i=1}^{d}$ and $\{q_i\}_{i=1}^{d}$ be marginal distributions corresponding to the best ICA approximation of p and q respectively. Then the change in mutual information between center and surround patches is

$$D\left(q \,||\, \prod_i q_i\right) - D\left(p \,||\, \prod_i p_i\right) = \int q \ln\left(\frac{q}{\prod_i q_i}\right) - \int p \ln\left(\frac{p}{\prod_i p_i}\right)$$

$$= \left[\sum_i H(q_i) - H(q)\right] - \left[\sum_i H(p_i) - H(p)\right]$$

$$= \left[H(p) - H(q)\right] + \sum_i \left[H(q_i) - H(p_i)\right]$$

$$= \Delta H(p; q) + \Delta H\left[(q_i); (p_i)\right] \tag{6.6}$$

where $H(p)$ is the entropy of distribution p, $\Delta H(p; q)$ is the entropy change between p and q, and $\Delta H[(p_i), (q_i)]$, is the cumulative entropy change between the MICA filter responses.

From (6.6) it follows that the CS NANS Index is also related to the entropy change between the center and surround patches. In particular, $\Delta H[(p_i), (q_i)]$ captures the entropy difference between the corresponding (MICA) filter responses of the center and surround patches, and $\Delta H(p; q)$ measures the overall entropy change between the center and surround patches.

In principle, the NANS Index can be implemented using MICA filters, ICA filters, or some other similar decomposition [76]. While MICA offers excellent advantages over ICA, both suffer from considerable computational complexity, since the MICA (or ICA) filters must be computed from every patch. Therefore, for the problem at hand, we have explored suboptimal approaches that deploy fixed filter sets. In particular, we derive a practical Gabor filter based NANS Index below—which we employ for fixation point selection.

Consider a bank of MxM Gabor filters that form a dyadic wavelet-like sampling of the frequency plane [148-150]. Let I_c and I_s be the center and the surround patches respectively. Then compute the $N=d$ dominant frequency channels as is done in [148-150] ($N < M^2$) corresponding to I_c, *viz.*, the N largest filter responses. Let $\{\phi_i\}_{i=1}^{N}$ be the resulting Gabor filters learned from the center patch. Assuming that I_s does not have the same dominant frequency channels as I_c (i.e. none of the dominant frequency channels of I_s correspond to that of I_c and vice versa), then

$$E\left[\left\| I_c - \sum_{i=1}^{N} < I_c, \phi_i > \phi_i \right\|^2\right] \le E\left[\left\| I_s - \sum_{i=1}^{N} < I_s, \phi_i > \phi_i \right\|^2\right]$$

Analyzing the both sides of the above expression separately:

$$LHS = E\left[< I_c - \sum_{i=1}^{d} < I_c, \phi_i > \phi_i, I_c - \sum_{i=1}^{d} < I_c, \phi_i > \phi_i > \right]$$

$$= E\left[\|I_c\|^2 \right] + \sum_{i=1}^{d}\sum_{j=1}^{d} E\left[< I_c, \phi_i >< I_c, \phi_j > \right] < \phi_i, \phi_j > - 2\sum_{i=1}^{d} E\left[|< I_c, \phi_i >|^2 \right]$$

$$= E\left[\|I_c\|^2 \right] + \|W \circ K_{\text{center}}\| - 2\,\text{trace}(K_{\text{center}})$$

$$= \left[\|W \circ K_{\text{center}}\| - \text{trace}(K_{\text{center}}) \right] + \Delta E_{\text{center}}$$

where

$$W = \left(w_{i,j} \right)_{i,j=1}^{d} = < \phi_i, \phi_j >_{i,j=1}^{d},$$

$$\Delta E_{\text{center}} = E\left[\|I_c\|^2 \right] - \text{trace}(K_{\text{center}})$$

K_{center} is the covariance matrix corresponding to the center patch, $\|\cdot\|$ is the L_1 matrix norm, and $"\circ"$ is the Hadamard (point-wise) product.

A similar expression holds for the RHS:

$$RHS = \left[\|W \circ K_{\text{surround}}\| - \text{trace}(K_{\text{surround}}) \right] + \Delta E_{\text{surround}}$$

where K_{surround} is the covariance matrix corresponding to the surround patch, and $\Delta E_{\text{surround}} = E\left[\|I_s\|^2 \right] - \text{trace}(K_{\text{surround}})$.

Then the Gabor-based CS NANS Index is

$$\eta = \left| 1 - \frac{E\left[\left\| I_S - \sum_{i=1}^{d} < I_S, \phi_i > \phi_i \right\|^2 \right]}{E\left[\left\| I_C - \sum_{i=1}^{d} < I_C, \phi_i > \phi_i \right\|^2 \right]} \right|$$

$$= \left| 1 - \frac{\|W \circ K_{\text{surround}}\| - \text{trace}(K_{\text{surround}}) + \Delta E_{\text{surround}}}{\|W \circ K_{\text{center}}\| - \text{trace}(K_{\text{center}}) + \Delta E_{\text{center}}} \right| \tag{6.7}$$

As with the MICA or ICA-based NANS Indices, the numerator and denominator of η consist of linear combinations of correlations between the various Gabor channels. Observe from (6.7) that computation of η does require computing correlation matrices corresponding to the center and surround patches; just the linear coding distortions of the center and surround patches.

As an example, Fig. 6.1(*a*) depicts an image of a fingerprint, which consists of highly oriented patterns with considerable local orientation and spatial frequency variance across the image. In places there are sudden changes in the pattern properties that can be ascribed to non-stationarity. Figure 6.1(*b*) shows the CS NANS Index map of the fingerprint. As can be seen the index values peak in a number of locations that are associated with local non-stationarity. It is important to note that overall fingerprint is, of course, highly non-stationary on a global scale, yet in most places is locally quite stationary except at certain singular locations. It is these features that the CS NANS Index is intended to highlight. As another example, Fig. 6.1(*c*) shows a natural image containing two primary substances: grass and water. The corresponding non-stationarity map is shown in Fig. 6.1(*d*). Although non-stationarities naturally occur at texture

boundaries, there are non-stationarities within some textures (such as grass) that are discovered by the CS NANS Index. In some other textures, such as the water texture, which is more spatially stationary (at least at this scale of analysis), there are few non-linearities detected. While the dominant non-stationarities in an image are typically induced by non-linear interactions between differing texture, there can be significant non-stationarities present within a given texture. Indeed, this type of non-stationarity may prove the most useful for characterizing, modeling, and recognizing textures.

We now formulate a greedy algorithm for determining the optimum texture-based fixations, which can be stated in terms of the following simple rule: The next optimum fixation point is simply the point in the image corresponding to the maximum non-stationarity (Gabor-based CS NANS Index).

An important consideration in implementing the above algorithm is the choice of the center-surround architecture to employ—apart from the image scale that we choose to analyze the non-stationary structure of images. In this chapter we analyze the non-stationary structure of the images at approximately the scale at which the foveola analyzes image regions assuming a viewing resolution of 1 arc minute per pixel.[2] The size of the center-surround windows described below reflects this choice of image scale. As discussed in Section 6.1, it is likely that the HVS uses a variety of different visual cues in order to determine visual fixations. The reason for this of course, given the hypothesized duality between NSS and low-level visual processes, is the considerable

[2] Since our goal in Section 4 is to compare fixation patterns obtained using non-stationarity index as a visual cue with human fixation patterns, this choice of pixel resolution sets a baseline with which performance with respect to HVS can be compared.

complexity of NSS—to overcome which, multiple low-level measurements must be made and subsequently integrated. The latter brings up the issue of combining visual cues.

6.4 Combining Texture and Contrast Fixation Features

Having developed optimal contrast and texture based fixations above, the question is what is the best way to combine these visual cues to yield optimal performance. The natural approach to this problem would be to formulate this as a joint optimization problem for extracting both contrast and textural information. However, we use a simpler approach for cue combination which is that of a simple alternation of contrast and texture based fixation patterns. As it turns out, this simple strategy performs remarkably well in modeling human fixations and in many cases outperforms both the contrast- and texture-based fixations performed separately.

Given a natural image, we compute optimal contrast-, texture- and simple combinations of texture-contrast fixations which are then compared to actual human fixations performed on those images. As a benchmark of performance, we compare the performance of the fixation algorithms to randomized fixation strategies and to other fixation-finding engines.

We obtained the human fixations from the DOVES database [151-152] wherein actual human fixations for grayscale images are recorded. The human fixations were recorded by using an SRI Generation V Dual Purkinje eye tracker. The stimuli (images) were displayed on a 21-inch, gamma corrected monitor at a distance of 134cm from the observer. The screen resolution corresponded to about 1 arc minute per pixel. Each image was displayed for 5 seconds in a fixed order for all observers. Observers were instructed

to free view each of the images as they desired. All observers commenced viewing the image stimuli from the center of the screen. The images were selected from the van Hateren database of natural images [13]. More details about the experimental set-up used to obtain fixations points can be found in [151-152]. We now detail the fixation algorithms described above.

The given image is first foveated in the center of the image such that the resolution is 1 arc-minute per pixel. This resolution is consistent with the set-up described in [67] used to obtain the human fixation results and is also consistent with the set-up used to obtain the contrast-based fixation patterns in Chapter 2. As in Chapter 2, 32x32 patches (0.53 deg) were used to compute the contrast image given the foveated luminance image. After multiple fixations, the image used for analysis is the contrast image corresponding to the Linear Scale Variant [69] image obtained by assigning a blur level to each pixel corresponding to the distance from the closest fixation point. From this contrast-of-LSV filtered image, an inference is made regarding the location of the next fixation point.

For the case of contrast-based fixations, the next fixation is determined by the MAP estimate using (6.3). For the case of texture-based fixations, the non-stationarity map was computed from above contrast-of-LSV filtered image. Given this non-stationarity map, as described in Section 6.3, the next fixation point is simply the point with the maximum non-stationarity measure. In order to speed up the computation of the fixation points, the NANS-based non-stationarity map is computed at sub-sampled pixel locations (with a sub-sampling factor of four) in the image following by interpolation to obtain dense non-stationarity maps.

In addition, for both the contrast- and texture- cases, since human fixations do not usually fall in the corners of the images, we neglect non-stationarity values within a width of 32 pixels from the borders of the image. This has the additional advantage of eliminating boundary effects since all the center-surround windows lie completely within the image. In order to avoid overly concentrated points of gaze, fixations are firstly forced not to fall within a foveal width of previous fixation points; and furthermore—following the procedure performed in [67]—the resulting selection map was also weighted using an inverted Gaussian mask centered on each selected fixation.

The random fixation patterns that are used to benchmark performance were generated in two different ways. In the first method, coordinates of the fixations were generated randomly according to a uniform distribution in each coordinate, with the constraint that the fixations should not lie within a foveal width of each other. We refer to fixation patterns generated in this way as *true random fixations*. Fixations patterns generated by the second method, which we call *HVS-based random fixations*, are obtained by shuffling the fixations recorded for an observer for particular (using the eye-tracking apparatus described in [67, 151-152]) with that of a different image. Thus the HVS-based fixation patterns simulate a random human observer whose fixations are not influenced by features of the underlying image, but otherwise captures all the statistics of human eye movements.

In our simulations we generate N_{fix} = 10 fixation points (beyond the first fixation point which is always at the center of the image) for each method (texture, contrast, texture-contrast and random fixation strategies). These fixation patterns must be compared with actual human fixation patterns obtained for the corresponding images.

One difficulty that arises in this process is that the lack of a definite order in the sequence of fixations (since we are not concerned here with the exact order of the fixations along the scanpaths, but only the locations in which they land in the image – although, of course, scanpath ordering is an exceedingly interesting question). Secondly, there is no guarantee that two fixations that are attracted by the same image structure may land within a fixed radius of each other. Further complicating the matter is that the number of fixation points being compared may not be equal; in particular the number of human fixation points that are available far exceed N_{fix}.

A simple and elegant solution to overcome these problems is to model the comparison process as a matching of probability distributions, by forming probability maps corresponding to the human fixations and the texture (or random) based fixations [146]. Probability maps were generated for true human fixations by placing a Gaussian of one foveal width at each of the fixation points followed by normalization (to form a probability density). Corresponding to each of the algorithm generated fixation maps (i.e. for texture-contrast, random, GAFFE etc), we generate *two* probability maps which are obtained, respectively, by placing Gaussian maps of one and two foveal widths on each of the fixation points. As we demonstrate below, evaluation of the performance of the various fixation algorithms with respect to these two probability maps gives us a more complete picture of the relative performances of the various algorithms.

A visual fixation algorithm may be viewed as succeeding as a predictor of human visual fixations strategies, if the spatial distributions or probability maps of computed fixations over a wide range of images agrees with the probability maps of human fixations on the same images. The fixation algorithms proposed here are based on

acquiring as much information (contrast and non-stationarity) as possible from an image via a fixation strategy.

Given the generated probability maps, the performance of a fixation algorithm is determined by the distance of one probability map to the other—in particular, the closer the two are, the better is the performance. A natural choice of measuring closeness of two probability distributions is the information-theoretic KLD. However since the KLD between two distributions is non-commutative, we seek symmetric extensions of the KLD to compare two probability distributions. We consider two such symmetric extensions, although many are possible.

Given two probability distributions p and q, one measure of symmetric KLD originally proposed in [154] is the average of the forward and reverse KLDs:

$$D_{\text{ave}}(p; q) = \frac{D(p \| q) + D(q \| p)}{2}$$

where for the sake of the following discussions we shall refer to $D(p \| q)$ as *forward KLD* and $D(q \| p)$ as *reverse KLD*. The second version of the symmetric KLD that we employ is the harmonic mean of the forward and reverse KLDs [155]:

$$D_{\text{harmonic}}(p; q) = \left(\frac{1}{D(p \| q)} + \frac{1}{D(q \| p)} \right)^{-1}$$

In this chapter, we examine the performance of the various fixation algorithms with respect to D_{ave} and D_{harmonic}.

Given probability maps p and q, the average KLD $D_{\text{ave}}(p, q)$ between them will be small (indicating a good match) only if *both* the forward and reverse KLDs are small. The same is true of $D_{\max}$. By contrast, $D_{\text{harmonic}}(p, q)$ will be small if either the forward or

reverse KLDs is small. Thus, if either D_{ave} or $D_{\max}$ is made small, then a more certain good match between the distributions p and q may be interpreted to exist than if D_{harmonic} is made small. However, any discrepancies between forward and reverse KLDs bears examination.

Next we examine the relative performances of the different fixations strategies. Figures 6.2-6.11 depict side-by-side comparisons of texture-contrast, GAFFE and human fixation patterns. The most immediate difference between the texture-contrast fixation patterns and the GAFFE fixation maps are that the texture-contrast fixations are more spread out across the images and generally more deeply intersect the effective domains of the human fixations. The GAFFE fixations generally cluster in tighter groupings. However, closer examination of the images reveals that the more widely spread texture-contrast fixations appear to fall at or near points of possible high visual saliency, *viz.*, near changes in texture, borders of shadow and light/dark, isolated contrast features, and so on. However, firmer conclusions may only be obtained by statistical comparisons with human fixations.

Tables 6.1-A and 6.1-B show the quantitative performance of the various fixation strategies when placing Gaussian windows of unit foveal width at the fixation locations generated by the algorithms. Table 6.1-A quantifies performance with respect to D_{ave}, while Table 6.1-B quantifies performance with respect to D_{harmonic}. We observe that whereas the performance of texture-contrast fixations exceed all other fixation strategies with respect to D_{ave}, GAFFE exceeds all other fixation strategies with respect to D_{harmonic}. These results taken together highlight the disparity between the forward and reverse KLD performance of GAFFE. This can be qualitatively understood as follows.

First, as observed before, the GAFFE fixations are not as spread out as the texture-contrast fixations. While this might be expected to occur at times owing to the peculiarities of some images, more generally a broader coverage of image space would be desired matching human behavior. Let p be the probability distribution associated with the human fixation patterns—an example of which is shown in Fig. 6.12(*a*) (corresponding to image #232); and let q be the probability distribution associated with the corresponding GAFFE fixation pattern—Fig. 6.12(*b*). It is clear that $D(p\|q)$ will be high for GAFFE, since many significant image regions heavily weighted by p are not matched by q.

Conversely, since the values of D_{harmonic} are quite low for GAFFE, this suggests that GAFFE performs well in matching human fixations over the "domain" of its attention. This observation suggests that complementary measures of the KLD, such as D_{ave} and D_{harmonic} (or perhaps the forward and reverse KLDs) can be useful for assessing the efficacy of fixation algorithms as comparatives to human fixations, rather than adhering to a single definition of KLD.

We also varied the widths of the interpolating Gaussians to better cover the image plane. Tables 6.2-A and 6.2-B show the performance of the various algorithms using interpolating Gaussians of two foveal widths (retaining unit-width Gaussians for the denser human fixations), while Fig. 6.12 gives a comparison of interpolation of fixations using unit- and twice-foveal-width Gaussians. Tables 6.2-A and 6.2-B show that the texture-contrast fixations outperform all other strategies with respect to both D_{ave} and D_{harmonic}. This is consistent with visual inspection of Figs. 6.2-6.11.

For the purpose of comparison, we have also included comparisons with Itti's fixation algorithm [64, 156], which was also studied in [67]. Itti's algorithm, which uses a broader suite of features than GAFFE or those used here, yields a performance that is competitive with GAFFE using the KLD measures, although fell short of the performance of GAFFE using a correlation measure [67].

The results here suggest that points of elevated contrast information or texture information supply features that appear to coincide with images features that draw visual attention. The types of features used by GAFFE and the Itti algorithm also fit this description, and indeed, there is no doubt a shared redundancy between the algorithms. It is likely that coupling the methods produced here with other algorithms, such as GAFFE, might generate even better results. However, the methods developed here more complex and highly nonlinear, it is unlikely that a simple linear weighting approach as used by GAFFE would be easy to implement.

When assessing the comparative performance of contrast- and texture- based fixations, we observe that for some images (such as image#245) the performance of the texture approach exceeds that of the contrast approach, and vice versa (such as in for image#54). While on average, both contrast and texture based fixations outperform random-based fixation strategies, we observe that the deviation of performance of contrast fixations from average is lower than for texture fixations. The combined texture-contrast fixations, however, outperforms both texture- and contrast- approaches performed separately. We finally point out that for the fixation strategies presented in this chapter there are some cases where they under-perform random fixations. The likely reason for this is that exploring one type of image feature, such as contrast or texture, can

potentially exclude it from exploring image regions where a complementary image cue dominates. It is for this reason that the HVS likely uses a combination of many different cues (both high and low level cues) when determining visual fixations of which only contrast and texture are explored in this chapter.

Figures 6.13-6.16 graphically summarize the above observations. The error bars in these figures indicate the positive and negative standard deviations of the respective quantities under consideration.

Although the results of this chapter indicate that contrast and texture are significant image features that determine low-level visual fixations, they are hardly the only cues employed by the HVS even amongst grayscale luminance image features (leaving out color, 3-D, motion, etc.). We envision that algorithms of this type will prove increasingly useful in the future, as large-format display systems create vast visual data streams, as databases of visual data increase, demanding greater increased content search efficiencies, and as mobile-camera robotic systems develop.

Questions as to the how the HVS actually computes and utilizes non-stationary natural image structure remain open. Although a Gabor-based implementation is certainly biologically plausible, and though the approach presented here is natural and relatively simple, the actual mechanisms employed by the HVS for non-stationarity handling in human perception are likely to be difficult to plumb, since non-stationarity is a subtle, indirect concept.

6.5 Discussion

The particular aspect of computational vision that we focus on in this chapter is fixation

selection. This can be viewed as a classic psychophysics problem wherein fixating vision system is treated as a black-box to which we feed visual stimuli (i.e. natural images) as input and observe the output behavior (visual fixations). Given the input and output data, the question is whether we can we gain system-level insights into the underlying mechanisms that system employs in determining visual fixations.

What we have uncovered in this dissertation is that an information-theoretic approach to this problem—along the lines of Barlow—can provide useful insights into these questions while also shedding light on the duality between NSS and low-level fixation processes in the HVS.

The existence of this hypothesized duality may well be because primitive visual systems—from which the HVS has evolved—adapted to the statistics of natural stimuli so as to develop near optimal responses that enabled the organism to efficiently gather information from the visual world in a survival-of-the-fittest setting. As one progresses to higher-level vision, however, such a duality might breakdown since we enter the progressively into the realm of cognition; indeed, it is not clear how meaningful and precise formulations within the optimal information processing framework can even be posed in such domains.

Nevertheless, we believe that a significant amount of work still remains in uncovering novel characterizations of NSS and exploring their possible relationships to visual fixations and other low-level visual processes such as contour grouping. To this end, the information-theoretic approach that we employed in this dissertation can potentially serve as a useful methodology for the formation of sharp hypotheses in ascertaining whether a given feature is actually a visual cue used by the HVS for the visual processes being

examined, and by possible extension, whether it might be used by computation algorithms seeking useful fixation or saliency points Two such visual cues (i.e. contrast and texture) have been established in this dissertation for the case of fixation selection in natural images.

Significant problems also remain in understanding the mechanisms underlying cue combinations. In this chapter we have shown that a simple interleaving between contrast and texture based fixation strategies yield robust performance across natural images relative to human performance. A rational survival strategy for an organism might be to continuously work at gaining as much general information as possible about the local environment until situations arise where the organism needs to be engaged in a particular task or until the organism detects a particularly significant object. This background of information could provide the grist for efficient performance in many of the organism's specific tasks. If this principle is correct, then a fixation selection mechanism based on maximizing the extraction of low-level visual information such as contrast or texture might work well as an automatic background or default mechanism that is overridden or modulated by more specific task demands or by particularly significant high-level content extracted during the fixations. Obviously, this is speculation, but it points to the real possibility that, in many cases, adequate models of human fixation patterns will require two or more very different fixation selection modules that interleave over time. With a deeper understanding into optimum cue combination mechanisms, we believe that hypotheses such as these can be tested and refined.

We envision that systematic progress along the lines described above will eventually lead to a unified information-theoretic understanding of low-level visual fixation

processes in the HVS, how they might be deployed in image and video processing systems, and generally yield insights into the deeper questions of visual understanding of natural scenes.

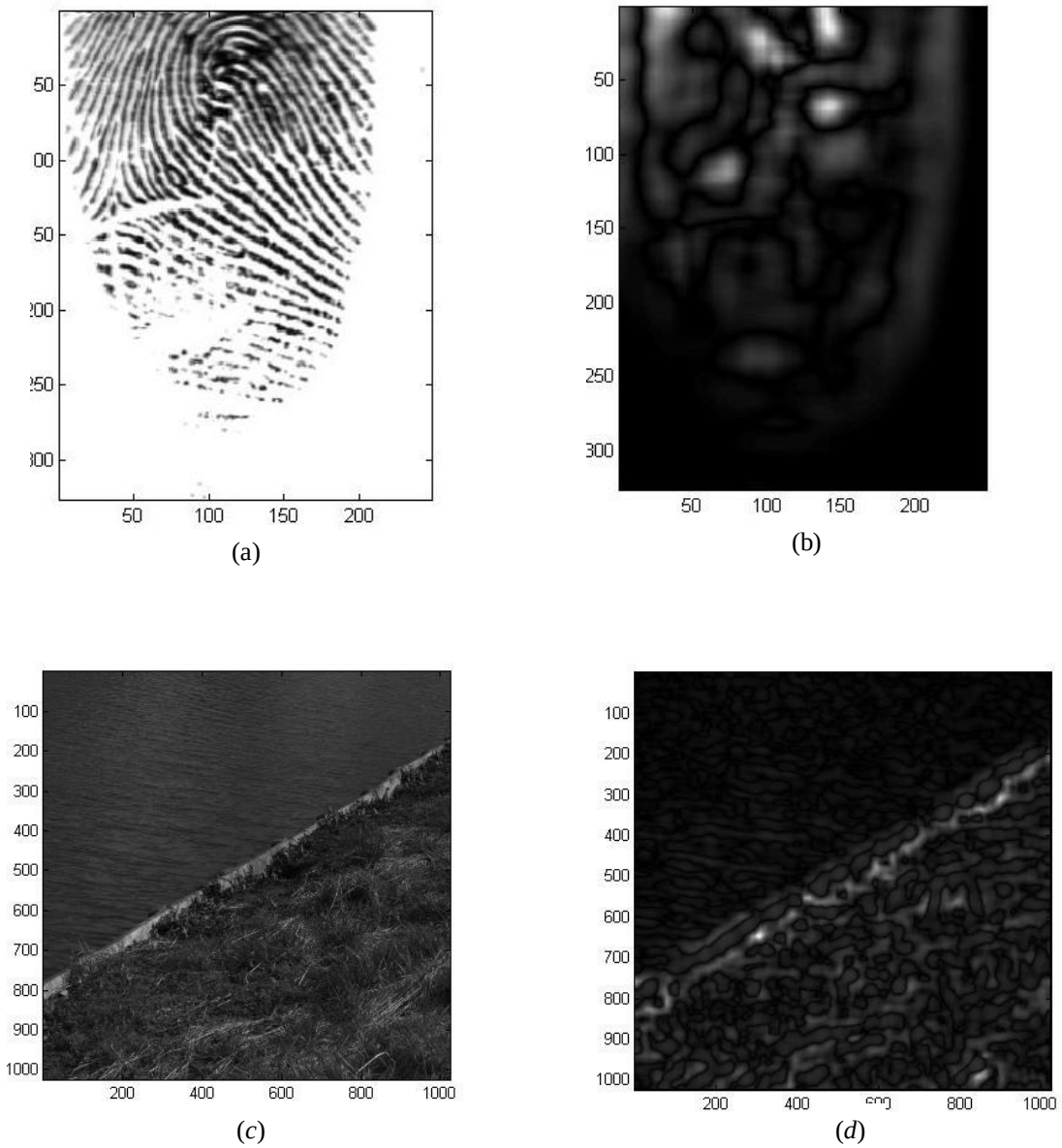
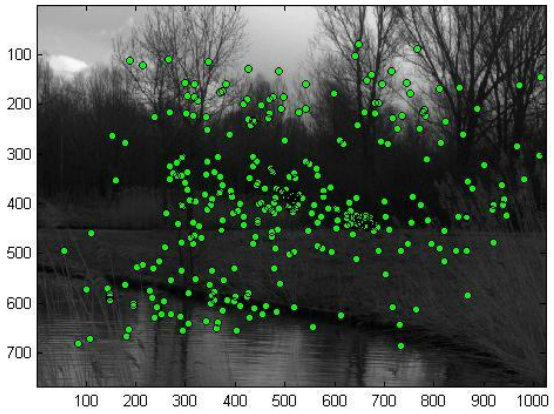
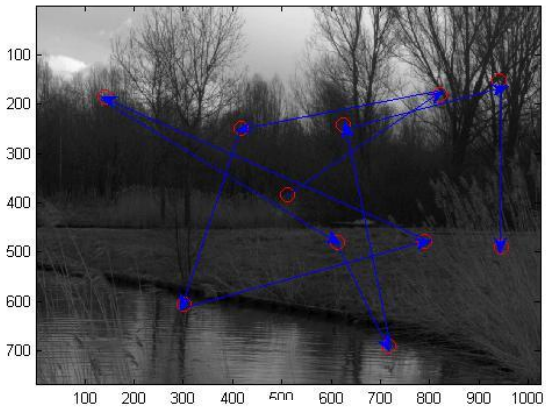


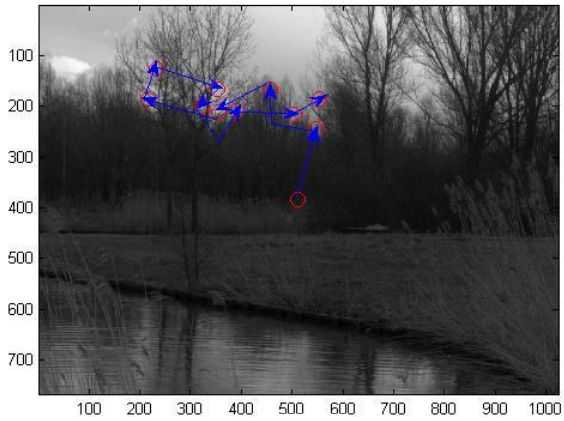
Fig. 6.1. Non-stationary analysis of a (*a*), (*b*) fingerprint image (*c*), (*d*) a natural image.



(a)

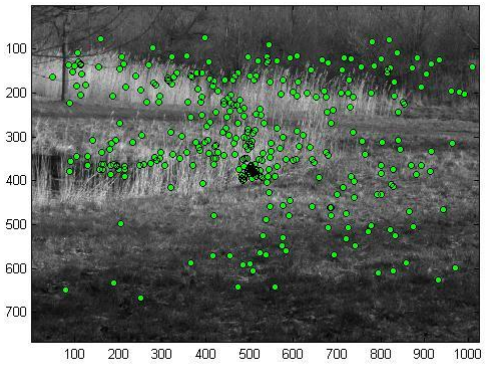


(b)

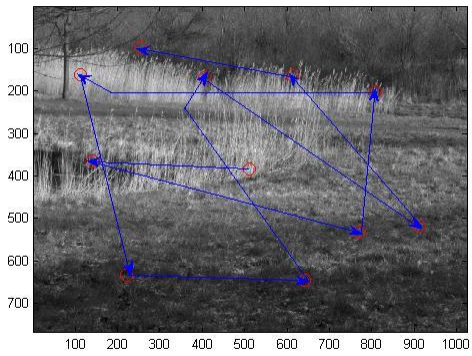


(c)

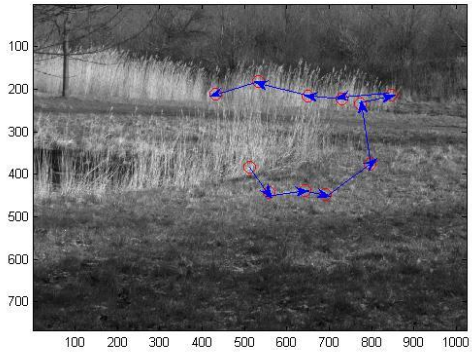
Fig. 6.2. Comparison of texture-contrast with human fixations on van Hateren image #245. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.



(a)

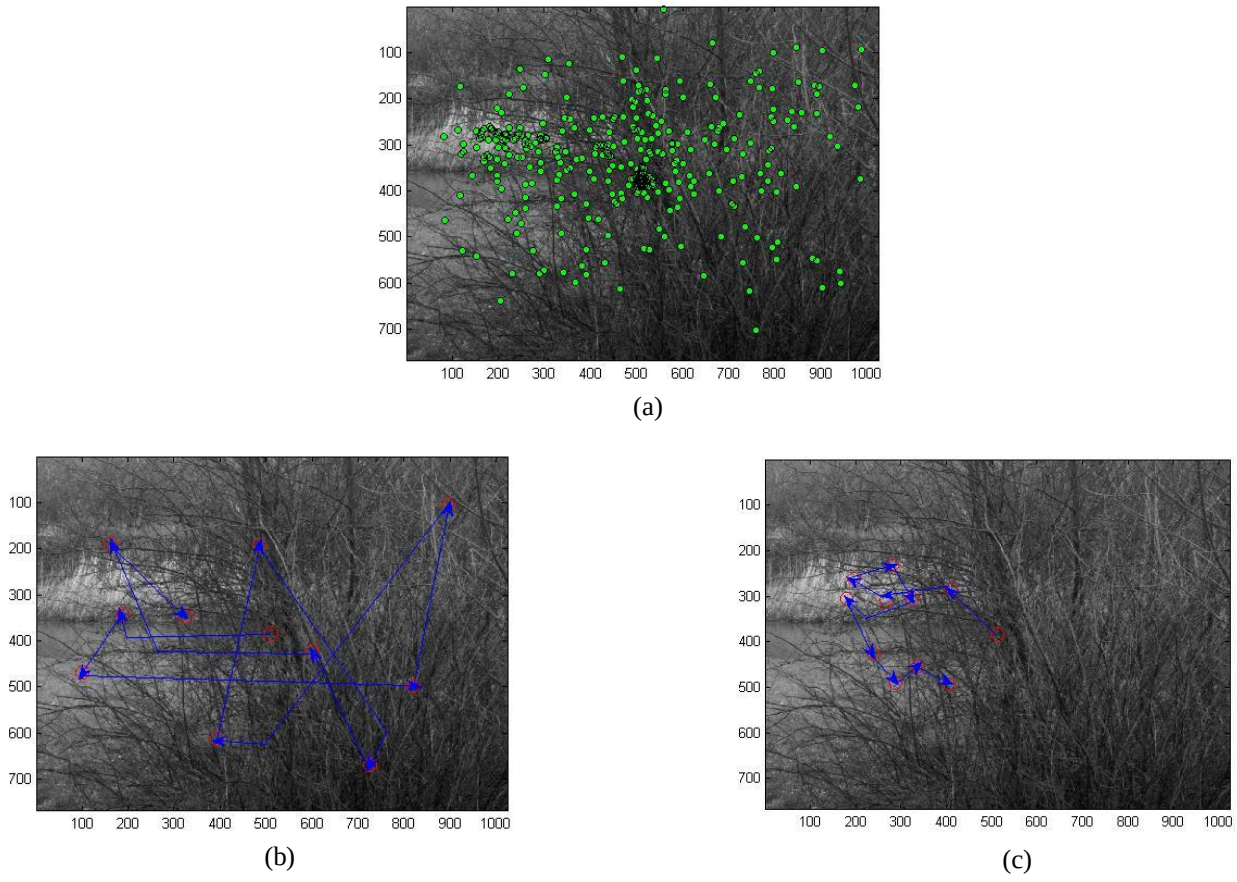


(b)



(c)

Fig. 6.3. Comparison of texture-contrast with human fixations on van Hateren image #37. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

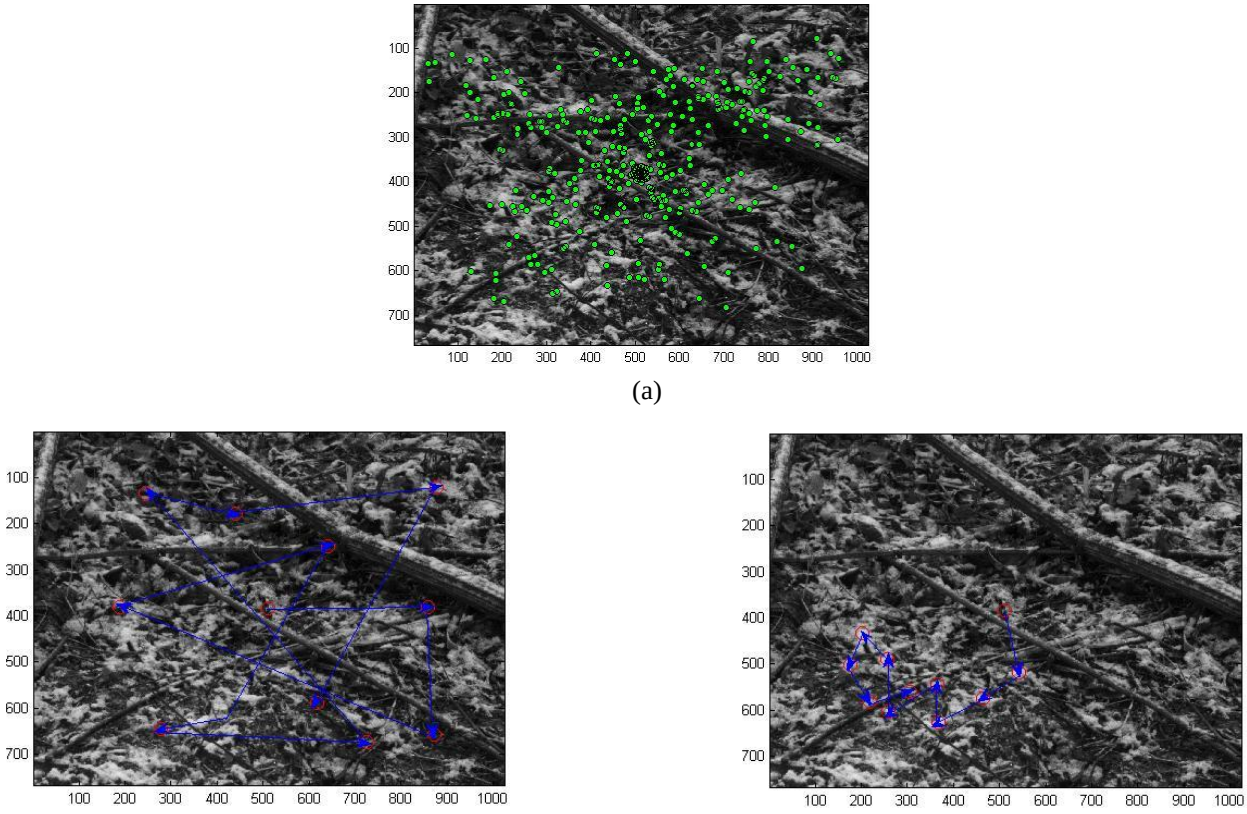


(a)

(b)

(c)

Fig. 6.4. Comparison of texture-contrast with human fixations on van Hateren image #122.
(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

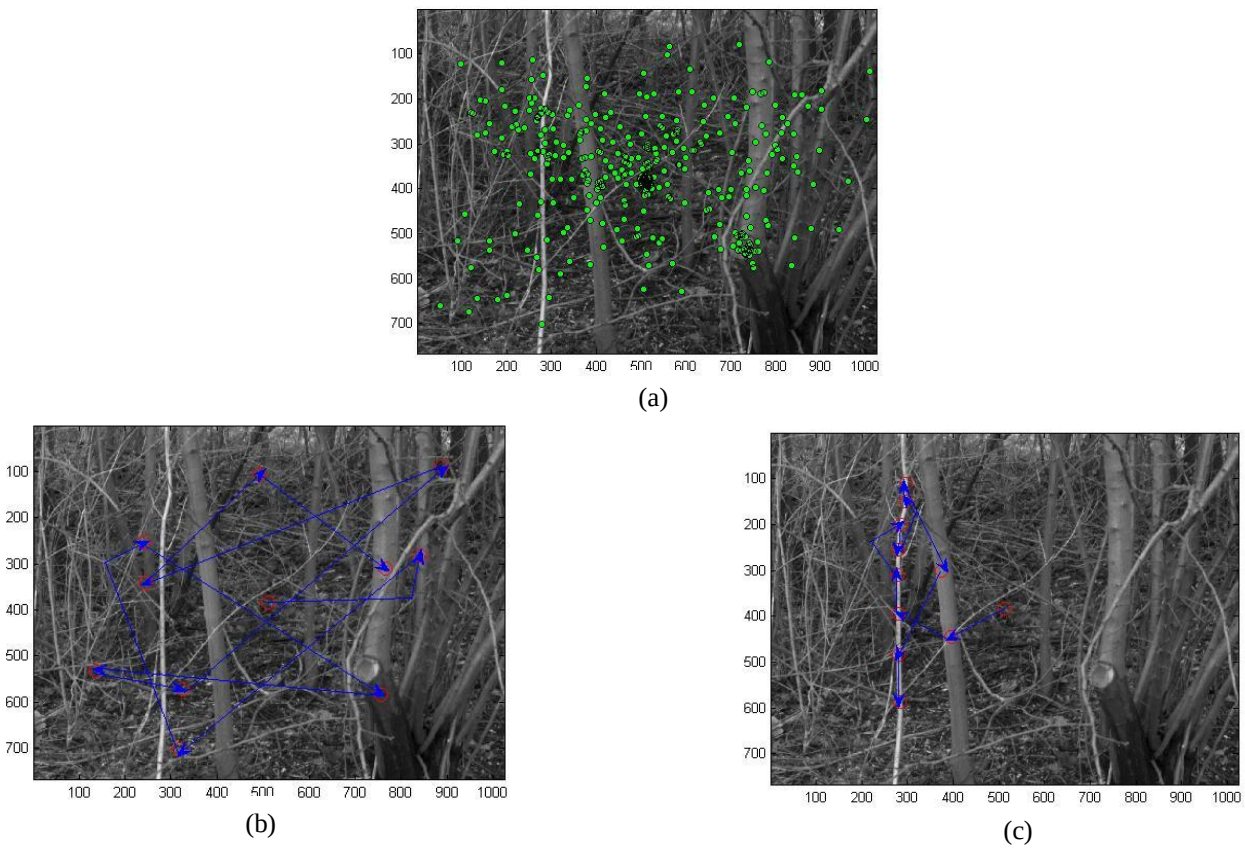


(a)

(b)

(c)

Fig. 6.5. Comparison of texture-contrast with human fixations on van Hateren image #34.
(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.



(a)

(b)

(c)

Fig. 6.6. Comparison of texture-contrast with human fixations on van Hateren image #161.
(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

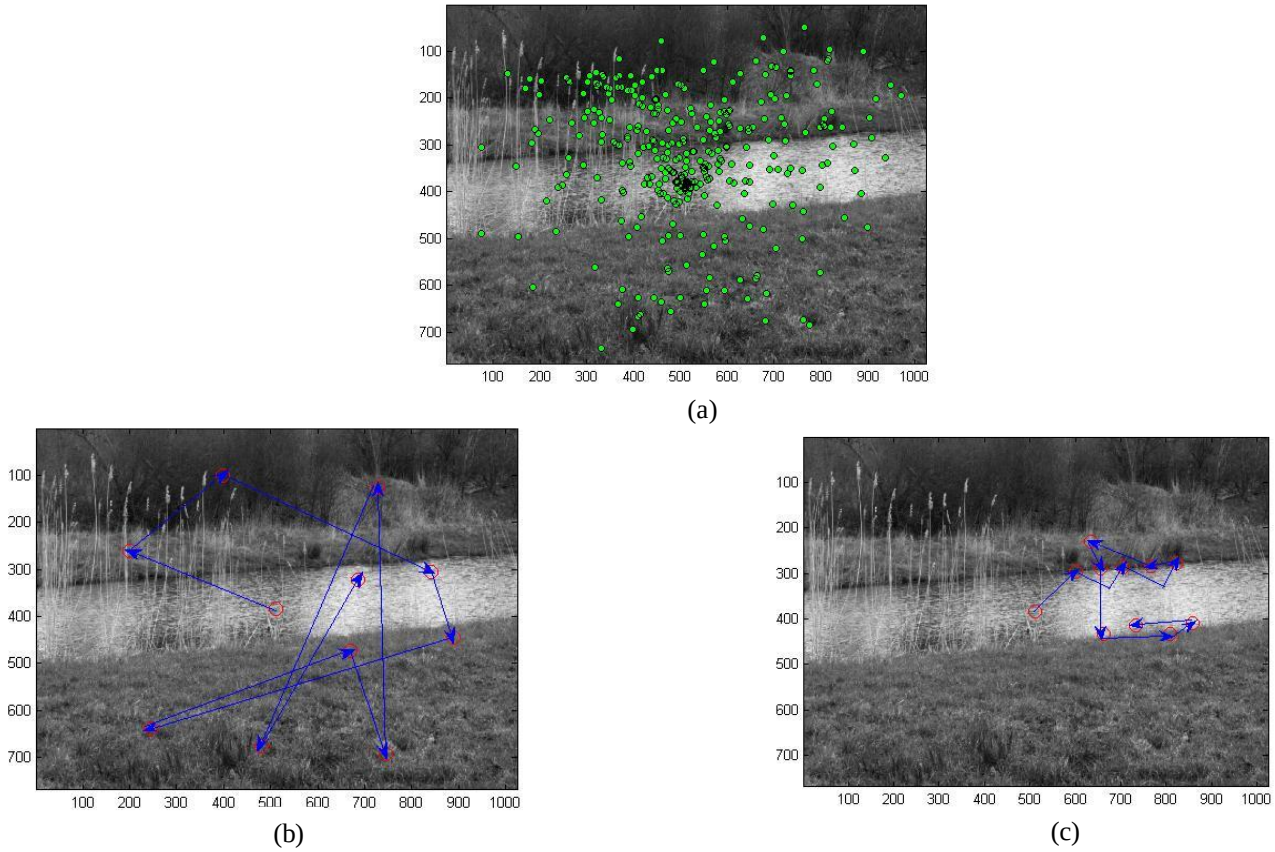


Fig. 6.7. Comparison of texture-contrast with human fixations on van Hateren image #232. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

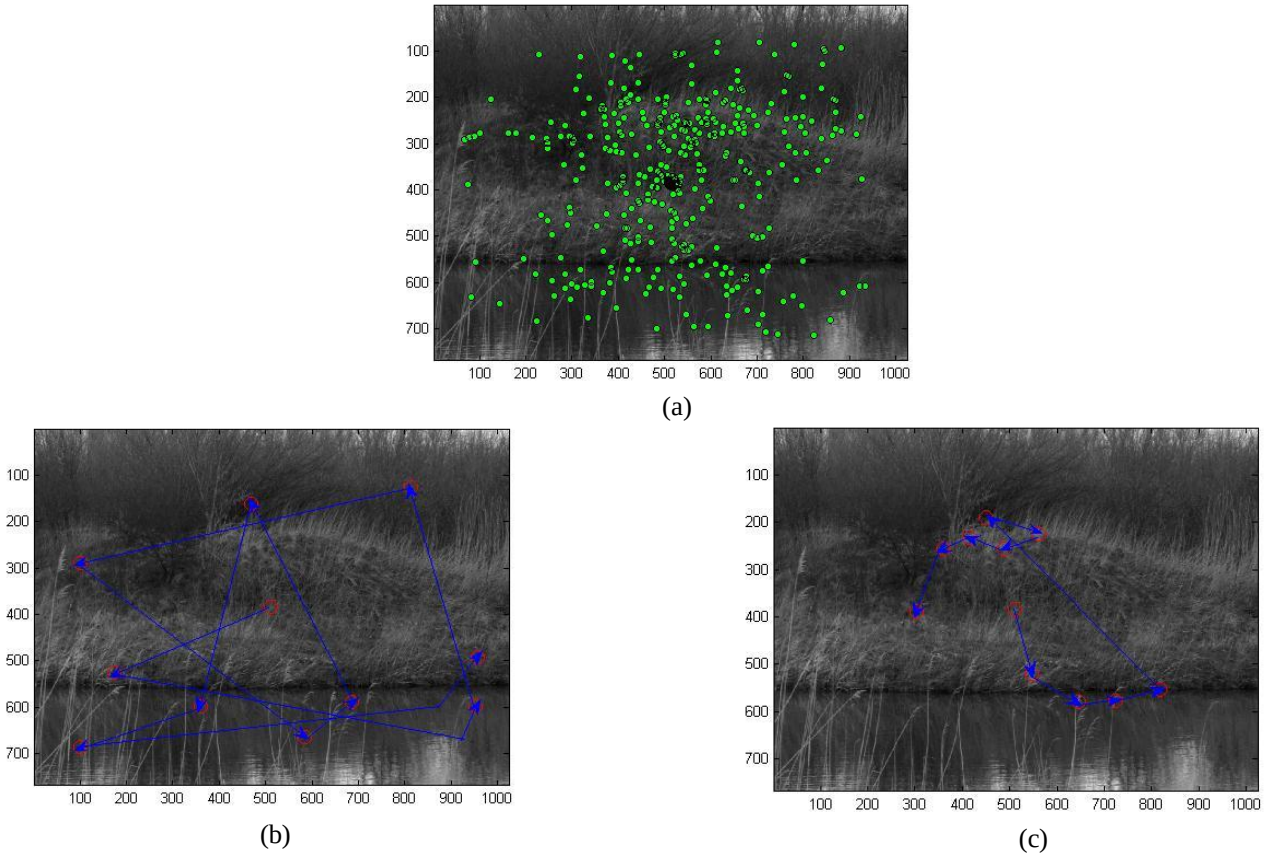
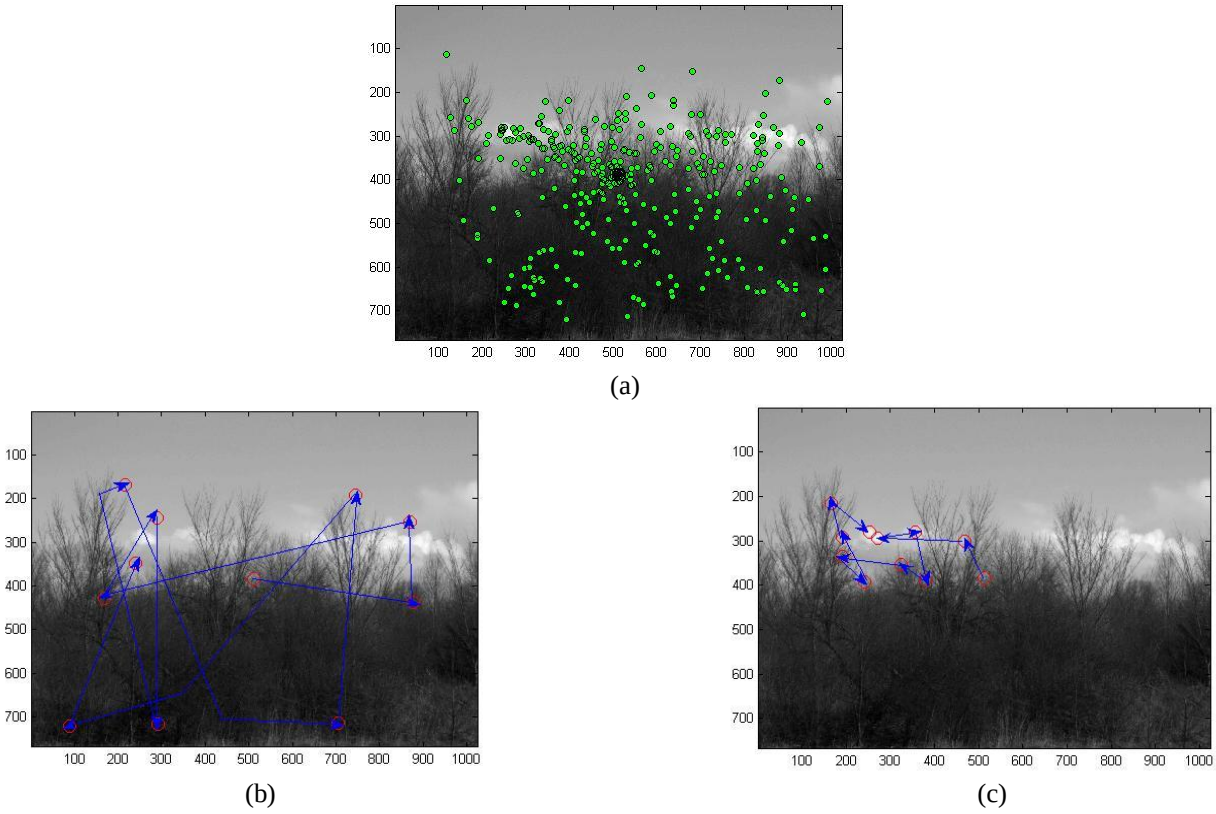


Fig. 6.8. Comparison of texture-contrast with human fixations on van Hateren image #146. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

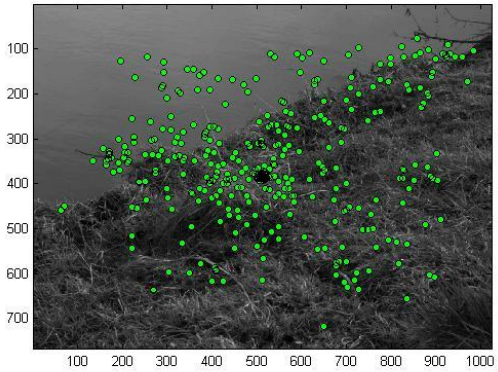


(a)

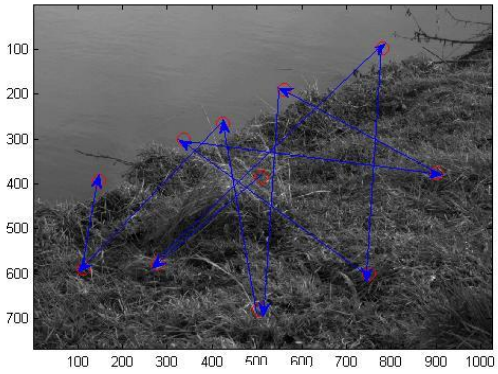
(b)

(c)

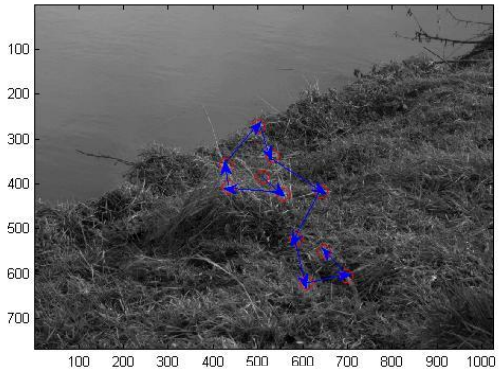
Fig. 6.9. Comparison of texture-contrast with human fixations on van Hateren image #54.
(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.



(a)

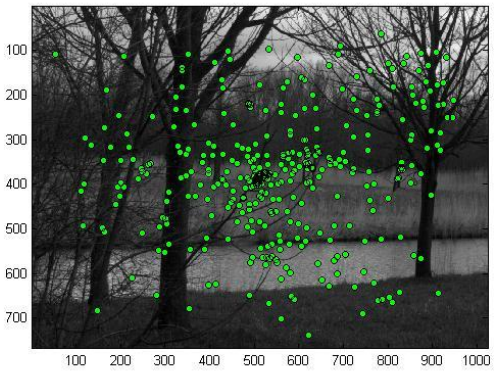


(b)

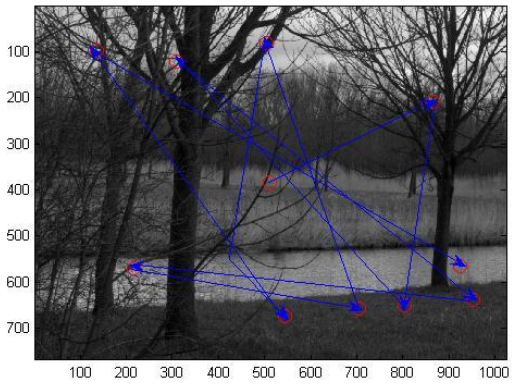


(c)

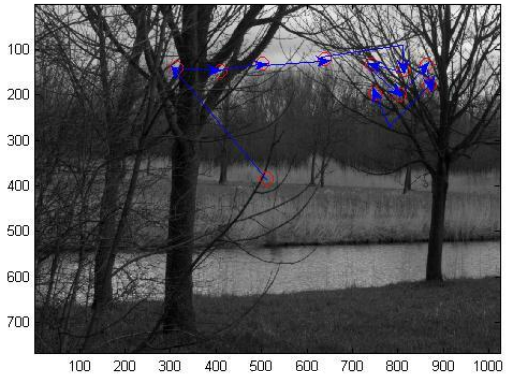
Fig. 6.10. Comparison of texture-contrast with human fixations on van Hateren image #353. (*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.



(a)



(b)



(c)

Fig. 11. Comparison of texture-contrast with human fixations on van Hateren image #190.
(*a*) Human fixations; (*b*) Texture-Contrast fixations; (c) Gaffe fixations.

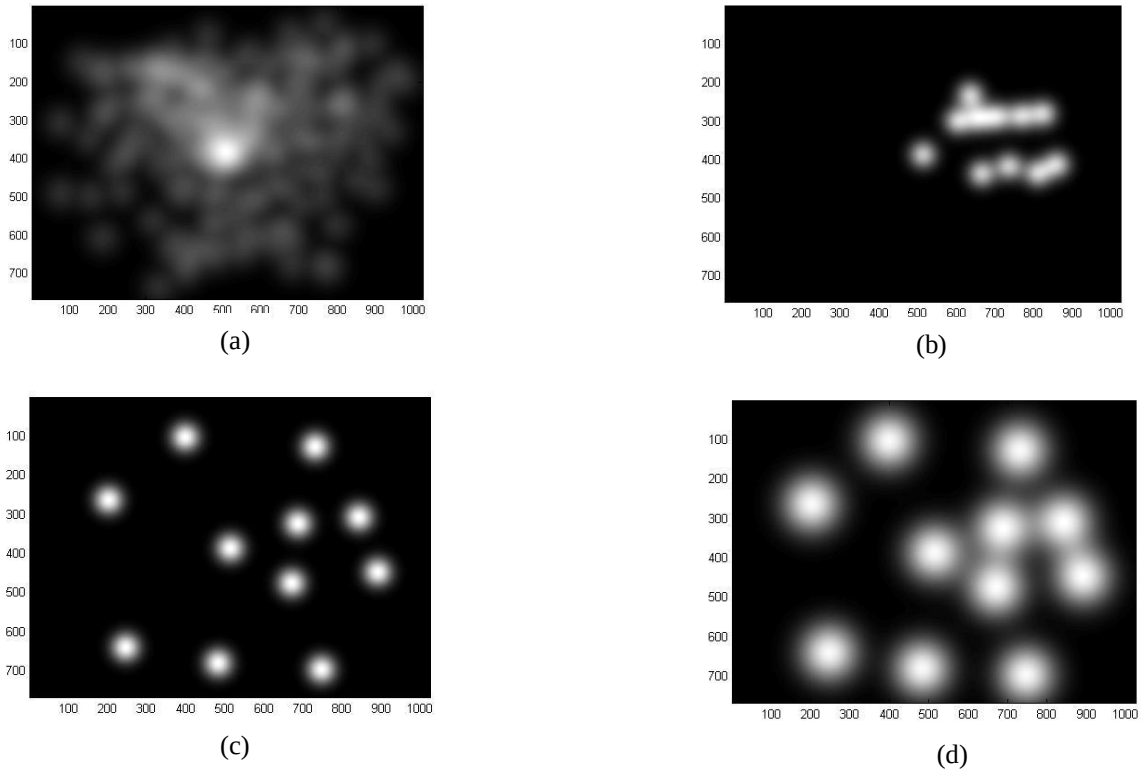


Fig. 12. Comparison of different probability maps corresponding to Image #232
(a) Human fixation induced probability map (unit-width Gaussians)
(b) GAFFE fixation map (unit-width Gaussians)
(c) Texture-contrast map (unit-width Gaussians)
(d) Texture-contrast map (twice-width Gaussians)

TABLE 6.1-A

Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{ave} and single-foveal width gaussian interpolation. The results shown are representative of the results obtained from a larger dataset in [152].

Image#	Texture fixations	Contrast fixations	Texture-contrast fixations	HVS-random fixations	True-random fixations	Gaffe fixations	Itti fixations
245	3.2941	4.7834	3.9558	5.4429	6.4711	8.4098	8.4368
37	3.6996	3.2614	3.4992	6.0413	6.2702	4.9942	4.2568
122	3.8604	4.3686	3.2237	4.7471	6.3973	4.6401	6.3947
34	4.0161	5.1686	4.0208	4.9509	6.0435	8.0543	3.8153
161	4.3834	5.4528	3.7423	5.0255	6.3193	6.0611	3.4627
232	3.9639	3.5132	3.7602	4.4218	5.9399	6.5396	5.3064
146	5.0571	4.7819	5.2546	4.4438	5.9636	4.4528	6.4843
54	5.2063	4.8891	5.9901	5.2035	6.4793	6.695	5.4482
353	5.52	5.4686	4.0821	5.0277	7.1128	6.3954	4.8286
190	7.4941	4.2009	5.8459	4.8889	5.3987	6.8553	7.291
Average(KLD)	**4.6495**	**4.58885**	**4.33747**	**5.01934**	**6.23957**	**6.30976**	**5.57248**

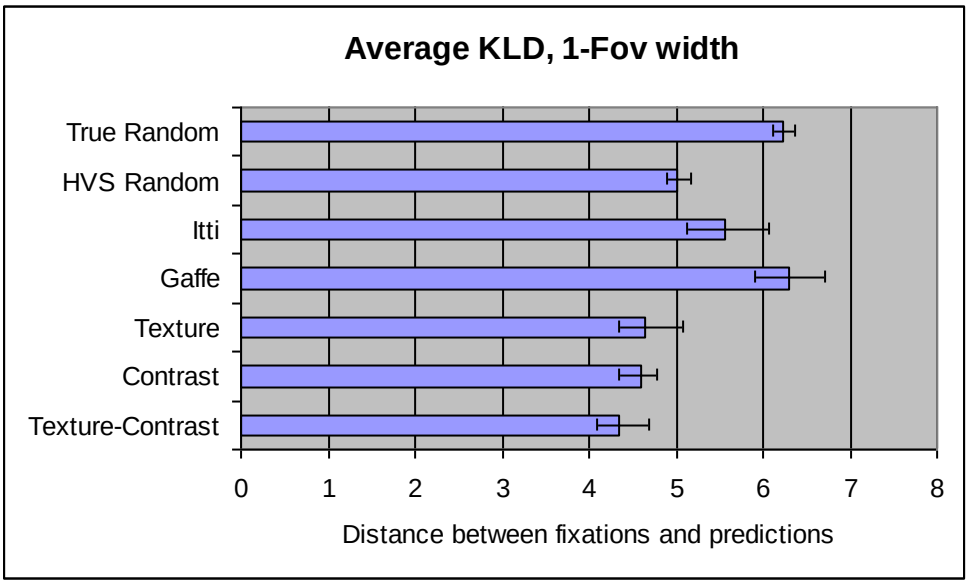


Fig. 6.13. Average KLD, 1-Foveal Width. Error bars indicate standard deviations.

TABLE 6.1-B

Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{harmonic} and single-foveal width gaussian interpolation. The results shown are representative of the results obtained from a larger dataset in [152].

Image#	Texture fixations	Contrast fixations	Texture-contrast fixations	HVS-random fixations	True-random fixations	Gaffe fixations	Itti fixations
245	1.2995	1.8443	1.5941	1.7089	2.6977	1.8856	2.2482
37	1.519	1.3005	1.4535	1.5324	2.405	1.2396	1.3389
122	1.6619	2.0772	1.2964	1.4703	2.8322	1.0397	1.3334
34	1.6638	2.216	1.7875	1.6013	2.5326	1.881	1.4736
161	2.0679	2.562	1.6449	1.6539	2.6947	1.4027	1.2422
232	1.5912	1.5625	1.5527	1.3377	2.5966	1.5096	1.7015
146	1.6057	1.9086	2.1186	1.395	2.3348	1.1509	2.2946
54	2.1708	2.3974	2.8984	1.7466	2.9229	1.3367	1.6802
353	2.1891	2.5955	1.9102	1.6421	3.2241	1.3161	1.765
190	2.3312	1.6948	2.1946	1.393	2.1317	1.7449	2.2168
Average(KLD)	**1.81001**	**2.01588**	**1.84509**	**1.54812**	**2.63723**	**1.45068**	**1.72944**

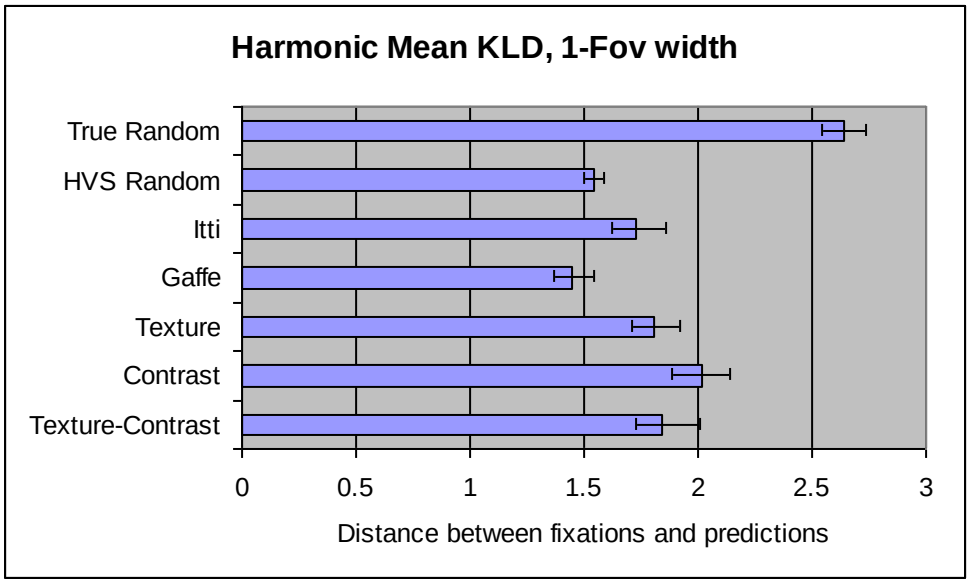


Fig. 6.14. Harmonic Mean KLD, 1-Foveal Width. Error bars indicate standard deviations.

TABLE 6.2-A

Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{ave} and twice-foveal width gaussian interpolation. The results shown are representative of the results obtained from a larger dataset in [152].

Image#	Texture fixations	Contrast fixations	Texture-contrast fixations	HVS-random fixations	True-random fixations	Gaffe fixations	Itti fixations
245	1.1142	1.5946	1.3281	1.986	2.4043	4.3686	4.0064
37	1.2017	1.0786	1.1193	1.8352	2.581	2.1097	1.1643
122	1.3586	1.8018	1.0174	1.8417	2.4951	2.4623	3.1594
34	1.4086	1.9697	1.4855	1.9289	2.4192	4.0686	1.1494
161	1.8018	2.4158	1.359	1.6706	2.5388	2.9094	1.0616
232	1.3814	1.3896	1.2472	1.475	2.2375	2.8485	1.8572
146	1.6432	1.4949	1.7977	1.805	2.1517	1.4879	2.7825
54	2.1357	1.9947	2.4927	2.4978	2.6111	3.7822	2.3936
353	2.6677	2.4809	1.5632	1.9788	2.6573	2.7658	1.6131
190	3.5434	1.3738	2.1103	1.8632	1.8601	3.005	3.0037
Average(KLD)	**1.82563**	**1.75944**	**1.55204**	**1.88822**	**2.39561**	**2.9808**	**2.21912**

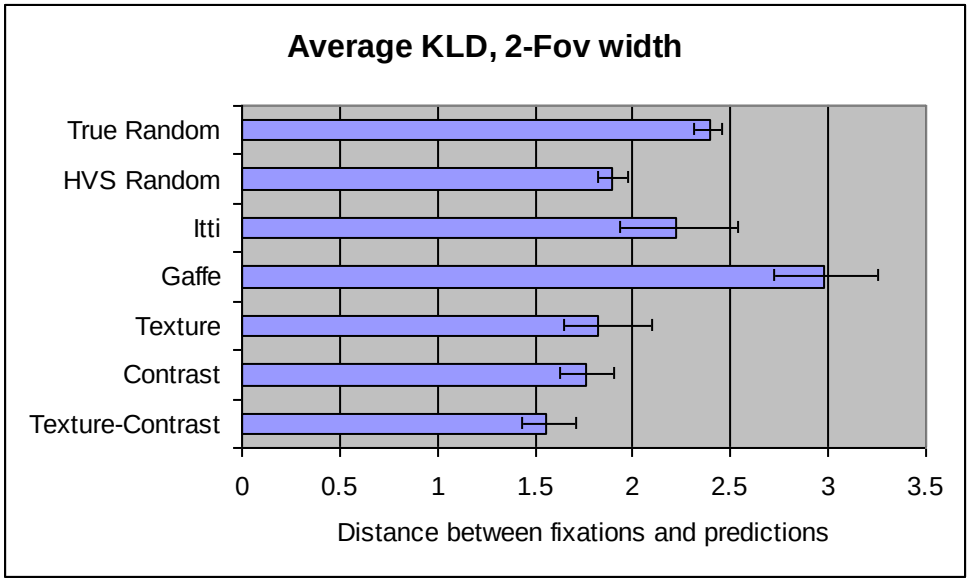


Fig. 6.15. Average KLD, 2-Foveal Width. Error bars indicate standard deviations.

TABLE 6.2-B

Summary of average performance (relative to true human fixations) of texture, contrast, and texture-contrast fixations as compared with true random, HVS-random, GAFFE and Itti fixations using D_{harmonic} and twice-foveal width gaussian interpolation. The results shown are representative of the results obtained from a larger dataset in [152].

Image#	Texture fixations	Contrast fixations	Texture-contrast fixations	HVS-random fixations	True-random fixations	Gaffe fixations	Itti fixations
245	0.5557	0.7965	0.6583	1.1572	1.1852	1.3731	1.5338
37	0.5926	0.5318	0.5502	1.2405	1.2789	0.6575	0.5515
122	0.6631	0.7919	0.5054	1.2638	1.2448	0.6615	0.9647
34	0.6856	0.9403	0.6943	1.2785	1.1771	1.3855	0.5746
161	0.8001	1.0474	0.6439	1.4618	1.2673	0.9446	0.53
232	0.6868	0.6395	0.6202	1.4893	1.1088	0.9513	0.8784
146	0.8021	0.7461	0.8896	1.2851	1.0741	0.5888	1.3287
54	1.0676	0.83	1.078	0.8332	1.2706	1.0964	1.0405
353	1.3164	1.1148	0.6907	1.1686	1.325	0.8647	0.8041
190	1.5652	0.6817	1.0519	1.3382	0.9298	1.1489	1.3248
Average(KLD)	**0.87352**	**0.812**	**0.73825**	**1.25162**	**1.18616**	**0.96723**	**0.95311**

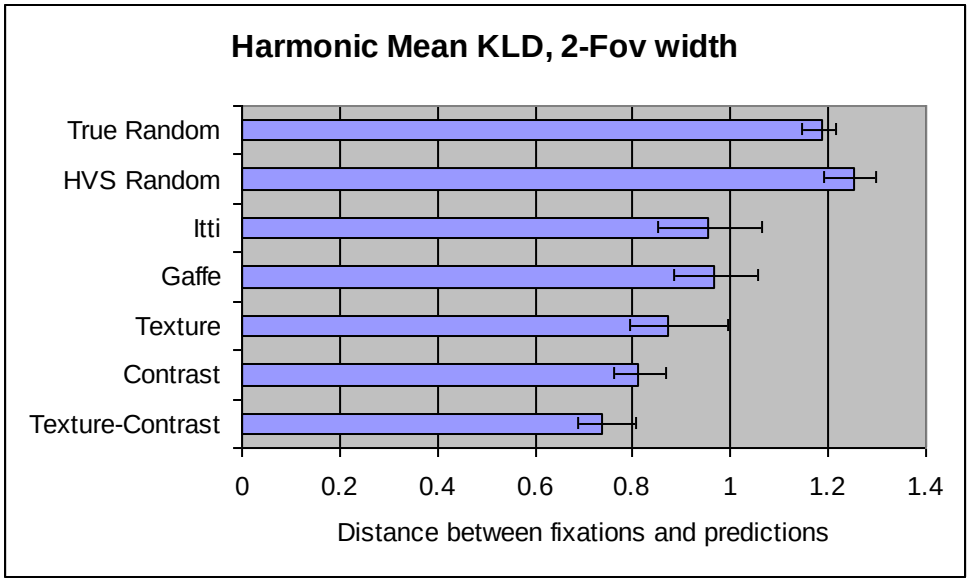


Fig. 6.16. Harmonic Mean KLD, 2-Foveal Width. Error bars indicate standard deviations

Chapter 7

Contributions and Future Work

7.1 Contributions

The end goal of the discipline of computational vision is, of course, a complete understanding of the computational aspects of the phenomenon of vision. As we discussed in Chapter 1, tremendous guidance to the sub-field of visual perception is offered by Barlow's hypothesis which postulates that the early stages of the visual system are designed to optimally extract relevant visual information from the surroundings. This induces the dual goal of identifying important aspects of NSS and discovering whether there is a corresponding implication to early visual processing. Impressive progress been made thus far in this line of work as described in Chapter 1.

Our overarching goal at the outset of this dissertation was to determine whether Barlow's hypothesis can be extended to encompass low-level fixation processes. The answer furnished by this dissertation, for the case of contrast and textural information, is in the affirmative. This provides a strong impetus to extend this line of work even further with the goal of developing a unified information-theoretic formulation of low-level visual search processes. Before we discuss future extensions of our work in the next section, we briefly summarize the major contributions of this dissertation.

Since foveation is an important aspect early visual processing performed by the HVS, in Chapter 2 we studied foveated signals in the more general setting of LSV signal processing. In particular, we established tight bounds of a point-wise linear approximation of linearly post-processed LSV signals. These results provide precise

conditions under which an equivalent linear channel can account for the resulting image structure thus considerably simplifying the analysis of LSV filtered signals acquired by the HVS.

Given that contrast is by far the most important low-level image feature coded by the HVS, a natural question is whether there are systematic and efficient ways in which the HVS acquires contrast information from natural scenes. Chapter 3 explored this question in detail wherein we uncovered a strikingly simple characterization of contrast statistics which find direct application to the formulation of optimal contrast-based fixation strategies.

While contrast is a highly local image property, texture (which can be qualitatively described as a 'roughly' stationary spatial process) is a regional property requiring complicated probabilistic models of spatial structure. Thus one formulation of the problem of optimum extraction of textural information from an image can be stated in terms of a full-fledged texture-based segmentation of the image. On the other hand, from the point of view of low-level visual fixations, it becomes immediately clear that the problem can be alternatively posed in terms of fixating to points of maximum non-stationarity in the image. This theme is developed fully in the remainder of the dissertation culminating in Chapter 6 where we demonstrate not only the performance of contrast- and texture-based fixation strategies individually but also show how robust fixation patterns (in terms of matching actual human fixation patterns) can be obtained by a simple interleaving of these two fixation strategies. A more detailed investigation into optimal cue combination mechanisms is the subject of future work.

In Chapters 4 and 5 we developed the tools necessary to formulate the optimum texture-based fixation algorithm. In particular, our computational theory of non-stationarity measurement was fully developed in Chapter 5 which in turn depends on a novel characterization of image spatial structure called the MICA decomposition which was described in detail Chapter 4.

Thus we have developed a bottom-up computational theory of low-level visual search processes in the HVS in terms of optimal processing of contrast and textural information. These results not only point the way to a unified information-theoretic treatment of low-level fixation processes but also demonstrate the tremendous scope that the understanding of low-level visual information processing can afford in gaining an insight into the deeper questions of image structure and computational vision.

7.2 Future Work

In the preceding chapters we have discussed open problems that emerge from each of our specific contributions. Here we discuss some of the more important ones in terms of the broader implications to computational vision and image/signal processing.

A. Problems in Computational Vision

7.2.A-1 Systematic Examination of Optimal Cue Combination Mechanisms

As we have discussed in previous chapters, determining the mechanisms of cue combination in the HVS (and by implication, identification of important visual cues) is one of the central goals of low-level vision. An important way make inroads into this problem is to study fixation strategies that optimally combine two or more visual cues.

Having established contrast and textural cues in this dissertation, an important open problem is therefore to determine optimum cue combination strategies that we jointly optimize the extraction of contrast and textural information from natural scenes.

7.2.A-2 Relationship with Other Fixation Strategies

Although much of the behavior of the low-level fixation patterns of the HVS can be explained and predicted by the information theoretic framework which we have presented in this dissertation, from a scientific point of view it is still very useful to examine possible relationship with other fixation strategies such as GAFFE since the HVS could conceivably employ a suite of such different strategies when deploying visual fixations.

7.2.A-3 Incorporation of Color, 3-D and Motion

Color and visual cues indicating depth are other very significant low-level image features that likely draw fixations from HVS. To this end, it will likely be very fruitful to extend our information theoretic approach to encompass these visual cues.

Inference of depth can be made either from stereo measurements (i.e. by exploiting the fact that the HVS consists of two eyes which can be used infer depth [14]), or from 2-D grayscale images themselves. As a first step, 3-D cues due to grayscale images (i.e. without stereo) can be characterized—by exploiting, in part, the non-stationary characterization of the image (which we developed in Chapter 5)—and optimal fixation strategies for this special case can be subsequently determined. Once the 3-D inference mechanisms induced by grayscale cues have been factored out, the effects on fixations due to purely stereo effects can be assessed more accurately. Since stereo effects are

dominant only for short range distances, it is likely that these two will factors end up playing complementary roles.

In this dissertation we have been exclusively concerned with still images. A very natural extension to this is, of course, to incorporate the time factor i.e. to extend our understanding of low-level visual processes to the domain of natural video sequences.

Again, subsequent investigations into optimal cue combination mechanisms will also have to be conducted for all the above cases!

7.2.A-4 Inroads into Higher-level vision

A fascinating question is how object feature information can be incorporated into the visual search process. From an engineering point of view this would be an important step for creating practical algorithms for search objects in natural scenes. Even from our work so far we can see that detecting regions of high non-stationarities can greatly reduce the search for objects in natural scenes, since the regions in the image where an object is located will create high non-stationarities with respect to the background texture.

B. Problems in Signal/Image Processing

7.2.B-1 Extensions to Graphical models and Information Geometry

Our approach to probabilistically modeling spatial image structure is based the MICA decomposition which we described in detail in Chapter 4. In order to more completely characterize spatial stochastic processes, however, a natural extension of our approach would be to couple it with graphical models [157] wherein the spatial process is modeled

as a Markov graph such that random variables are associated with each of the nodes and where the edges between the nodes model dependences.

The problem that we propose in this regard is investigation into optimal algorithms for the construction of a Markov tree representations of the spatial stochastic process based on maximization of the sparsity of the connected nodes of the graph with the incorporation simple spatial constraints. Subsequent examination of the graph structure based on an information-geometric analysis [158] can shed significant light on the high-order structure of the spatial image process in a multi-scale fashion.

7.2.B-2 Relation to Divisive Normalization Contrast Measure

As explained in previous chapters, contrast is by far the most important low-level image feature coded by the HVS. There are however various characterizations of image contrast. In this dissertation we have exclusively dealt with RMS contrast from which we have obtained very useful results in terms of both statistical characterization and implications to fixation selection algorithms.

Another very commonly used contrast measure is the divisively normalized contrast [34-35] wherein the local image contrast is measure by the luminance at a given pixel normalized by the average luminance measured over a local window. Thus an important problem is to examine the statistics properties of this contrast measure for natural images and to examine its implications to optimum visual fixation algorithms.

7.2.B-3 Examination of MICA-based Sparse coding and Extensions of the MICA model

Consider a sparse coding problem involving the joint minimization of the MSE (i.e. mean-squared coding error with respect to $\{\phi_i\}_{i=1}^{d}$) and a sparsity term induced by $g(J)$. Is there an optimum basis set that is a solution to this problem?

In Chapter 4 we have considered the MICA model for the complete and under-complete cases. A first step to extending this to the over-complete case is to address the problem of parameter estimation of a mixture of MICA models. This would have the added benefit of enabling the analysis of data from multi-modal probability distributions.

Furthermore, given that the MICA model can be extended to arbitrary basis sets to form non-sparse probabilistic representations of natural image source data. The corresponding implications to the non-stationarity measurement in natural images can be numerically examined.

7.2.B-4 Rate-Distortion Coding Properties of MICA-based Non-linear Image Representation

The coding of an image in a sparse basis set such as MICA, implies that the representation is non-linear in the sense that the optimal sparse representation of the sum of two images will, in general, yield drastically different sparse codes as compared to the sparse codes of the original images. A rate-distortion characterization of MICA-based non-linear image representations will yield considerable insight into its efficacy in compressing natural image data.

Appendix A

Proofs of Chapter 2

Proof of Theorem 2.1 – We have that

$$|\varepsilon(\boldsymbol{x}_0)|^2 = \left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \left[\int_{\boldsymbol{R}^n} f(\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{a}) \cdot \frac{1}{[\sigma(\boldsymbol{x}_0)]^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x}_0)]\, d\boldsymbol{a} \right.\right.$$

$$\left.\left. - \int_{\boldsymbol{R}^n} f(\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{a}) \cdot \frac{1}{[\sigma(\boldsymbol{x}_0 - b)]^n} g[\boldsymbol{a}/\sigma(\boldsymbol{x}_0 - b)]\, d\boldsymbol{a} \right] d\boldsymbol{b} \right|^2 . \qquad \text{(A1)}$$

By making appropriate substitutions (A1) becomes:

$$|\varepsilon(\boldsymbol{x}_0)|^2 = \left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \int_{\boldsymbol{R}^n} \{ f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0)] - f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0 - \boldsymbol{b})] \} g(\boldsymbol{v})\, d\boldsymbol{v}\, d\boldsymbol{b} \right|^2 . \qquad \text{(A2)}$$

The term in the inner curly brackets can be evaluated using a first-order Taylor's approximation with explicit remainder [93, p. 203]:

$$f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0)] - f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0 - \boldsymbol{b})]$$

$$= \int_0^1 \sum_{i=1}^n v_i\, [\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)] \nabla f_i \{ \boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0) - s\boldsymbol{v}[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)] \} ds \qquad \text{(A3)}$$

where $\boldsymbol{v} = (v_1, \ldots, v_n)$. With this the squared error (A2) becomes:

$$|\varepsilon(\boldsymbol{x}_0)|^2 =$$

$$\left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) [\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)] \sum_{i=1}^n \int_0^1 \int_{\boldsymbol{R}^n} v_i g(\boldsymbol{v}) \nabla f_i \{ \boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0) - s\boldsymbol{v}[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)] \} d\boldsymbol{v}\, ds\, d\boldsymbol{b} \right|^2 \qquad \text{(A4)}$$

the innermost integral of which is bounded above by the Cauchy-Schwarz inequality:

$$\int_{\boldsymbol{R}^n} v_i g(\boldsymbol{v}) \nabla f_i \{\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0) - s\boldsymbol{v}[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\} d\boldsymbol{v}$$

$$\leq \sqrt{\Delta g_i} \cdot \sqrt{\int_{\boldsymbol{R}^n} \nabla f_i^{\,2} \left\{\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0) - s\boldsymbol{v}[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\right\} d\boldsymbol{v}} \qquad \text{(A5)}$$

where Δg_i is given in (7). By making the substitutions

$$\boldsymbol{r} = \boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\left\{\sigma(\boldsymbol{x}_0) + s[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\right\} \qquad \text{(A6)}$$

$$d\boldsymbol{r} = -\left\{\sigma(\boldsymbol{x}_0) + s[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\right\} d\boldsymbol{v} \,. \qquad \text{(A8)}$$

the term inside the radical in (A5) can be re-expressed as

$$\int_{\boldsymbol{R}^n} \nabla f_i^{\,2} \{\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0) - s\boldsymbol{v}[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\} d\boldsymbol{v} = \frac{1}{\sigma(\boldsymbol{x}_0) + s[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]} \delta f_i \qquad \text{(A9)}$$

where δf_i is given by (8). Using (A5)-(A9) yields the following bound on the squared error (A2):

$$|\varepsilon(\boldsymbol{x}_0)|^2 \leq$$

$$\sum_{i=1}^{n} \Delta g_i \, \delta f_i \left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b})[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)] \int_0^1 \frac{ds}{\left\{\sigma(\boldsymbol{x}_0) + s[\sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0)]\right\}^{n/2}} d\boldsymbol{b} \right|^2 \qquad \text{(A10)}$$

Making another substitution of variables in the innermost integral of (A10), letting $\Delta\boldsymbol{g} = (\Delta g_1, \ldots, \Delta g_n)^T$ and $\delta\boldsymbol{f} = (\delta f_1, \ldots, \delta f_n)^T$, then (A10) becomes

$$|\varepsilon(\boldsymbol{x}_0)|^2 \leq (\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f}) \left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \int_{\sigma(\boldsymbol{x}_0)}^{\sigma(\boldsymbol{x}_0 - \boldsymbol{b})} s^{-n/2} \, ds \, d\boldsymbol{b} \right|^2 . \qquad \text{(A11)}$$

Further evaluation of (A11) requires noting that the innermost integral yields different definite forms when $n = 2$ and $n \neq 2$. When $n \neq 2$,

$$\int_{a}^{b} s^{-n/2}\, ds = \left(\frac{2}{2-n}\right)\left[b^{1-(n/2)} - a^{1-(n/2)}\right] \tag{A12}$$

and when $n = 2$

$$\int_{a}^{b} s^{-1}\, ds = \ln(b/a)\,. \tag{A13}$$

For the case $n \neq 2$, the innermost integral becomes

$$\int_{\sigma(\boldsymbol{x}_0)}^{\sigma(\boldsymbol{x}_0-\boldsymbol{b})} s^{-n/2}\, ds = \left(\frac{2}{2-n}\right)\left\{[\sigma(\boldsymbol{x}_0)]^{1-(n/2)} - [\sigma(\boldsymbol{x}_0-\boldsymbol{b})]^{1-(n/2)}\right\} \tag{A14}$$

$$= \sum_{j=1}^{n} \int_{0}^{1} b_j \frac{\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})}{[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})]^{n/2}}\, ds\,, \tag{A15}$$

where $\nabla\sigma_i(\boldsymbol{x}) = \partial\sigma(\boldsymbol{x})/\partial x_i$, $i = 1$,.., n are the elements of the gradient vector of the scaling function. Equation (A15) follows by application of the closed-form of the first-order Taylor's approximation of the difference within curly brackets in (A14). Hence (A11) becomes ($n \neq 2$)

$$|\varepsilon(\boldsymbol{x}_0)|^2 \leq (\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f}) \left| \int_{0}^{1} \sum_{j=1}^{n} \int_{\boldsymbol{R}^n} b_j h(\boldsymbol{b}) \frac{\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})}{[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})]^{n/2}}\, d\boldsymbol{b}\, ds \right|^2 . \tag{A16}$$

For the case $n = 2$, as it turns out, the bound is the same, since

$$\int_{\sigma(\boldsymbol{x}_0)}^{\sigma(\boldsymbol{x}_0-\boldsymbol{b})} s^{-1}\, ds = \ln\left[\sigma(\boldsymbol{x}_0-\boldsymbol{b})/\sigma(\boldsymbol{x}_0)\right] = \sum_{j=1}^{n} \int_{0}^{1} b_j \frac{\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})}{\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})}\, ds \tag{A17}$$

and so it follows that the bound (A11) is given by (A16) for $n = 2$ as well. But this can be simplified even further by again applying the Cauchy-Schwarz inequality, this time to the innermost integral of (A16).

$$\int_{\boldsymbol{R}^n} b_j h(\boldsymbol{b}) \frac{\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})}{\left[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})\right]^{n/2}} d\boldsymbol{b} \le \sqrt{\Delta h_j} \cdot \sqrt{\int_{\boldsymbol{R}^n} \frac{\left|\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})\right|^2}{\left[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})\right]^n} d\boldsymbol{b}} \tag{A18}$$

where Δh_j, the directional energy variance of $h(\boldsymbol{x})$, is defined as in (7). Hence the squared error functional (A1) is further bounded as

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le (\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f}) \sum_{j=1}^{n} \Delta h_j \left| \int_0^1 \sqrt{\int_{\boldsymbol{R}^n} \frac{\left|\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})\right|^2}{\left[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})\right]^n} d\boldsymbol{b}}\, ds \right|^2 . \tag{A19}$$

The innermost integral of (A19) can also be simplified:

$$\int_{\boldsymbol{R}^n} \frac{\left|\nabla\sigma_j(\boldsymbol{x}_0 - s\boldsymbol{b})\right|^2}{\left[\sigma(\boldsymbol{x}_0 - s\boldsymbol{b})\right]^n} d\boldsymbol{b} = \frac{1}{s^n} \partial\sigma_j \tag{A20}$$

where $\partial\boldsymbol{\sigma} = (\partial\sigma_1, \ldots, \partial\sigma_n)^T$ is the vector of weighted derivative (Sobolev) norms given by (9). Also denoting $\Delta\boldsymbol{h} = (\Delta h_1, \ldots, \Delta h_n)^T$, the bound (A19) can be expressed

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le (\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f})\, (\Delta\boldsymbol{h} \bullet \partial\boldsymbol{\sigma}) \left| \int_0^1 s^{-n/2} ds \right|^2 \tag{A21}$$

However, $\int_0^1 s^{-n/2} ds$ evaluates to $2/(2-n)$ except when $n = 2$, in which case the integral does not converge (is infinite). Hence we finally have ($n \neq 2$)

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le \left(\frac{2}{2-n}\right)^2 (\Delta\boldsymbol{g} \bullet \delta\boldsymbol{f})\, (\Delta\boldsymbol{h} \bullet \partial\boldsymbol{\sigma}) \tag{A22}$$

which, after take the square root of each side, finishes the proof. ♣

Proof of Corollary 2.1 – We have from (A2) that

$$|\varepsilon(\boldsymbol{x}_0)|^2 = \left| \int_{\boldsymbol{R}^n} h(\boldsymbol{b}) \int_{\boldsymbol{R}^n} g(\boldsymbol{v}) \left\{ f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0)] - f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0 - \boldsymbol{b})] \right\} d\boldsymbol{v}\, d\boldsymbol{b} \right|^2 . \quad \text{(A23)}$$

By the Fundamental Theorem for line integrals we have that

$$f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0)] - f[\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0 - \boldsymbol{b})]$$

$$= \int_{\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0 - \boldsymbol{b})}^{\boldsymbol{x}_0 - \boldsymbol{b} - \boldsymbol{v}\sigma(\boldsymbol{x}_0)} \nabla \boldsymbol{f} \bullet d\boldsymbol{s} = \sum_{i=1}^{n} \int_{-v_i\sigma(\boldsymbol{x}_0 - \boldsymbol{b})}^{-v_i\sigma(\boldsymbol{x}_0)} \nabla_i f(\boldsymbol{x})\, dx_i . \quad \text{(A24)}$$

The squared error (A23) can be bounded as

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le \left[\sum_{i=1}^{n} \int_{\boldsymbol{R}^n} |h(\boldsymbol{b})| \int_{\boldsymbol{R}^n} |g(\boldsymbol{v})| \left| \int_{-v_i\sigma(\boldsymbol{x}_0 - \boldsymbol{b})}^{-v_i\sigma(\boldsymbol{x}_0)} \nabla_i f(\boldsymbol{x})\, dx_i \right| d\boldsymbol{v}\, d\boldsymbol{b} \right]^2 \quad \text{(A25)}$$

$$\le \delta f_{\max}^2 \left[\sum_{i=1}^{n} \int_{\boldsymbol{R}^n} |h(\boldsymbol{b})| \int_{\boldsymbol{R}^n} |v_i|\, |g(\boldsymbol{v})| \left| \sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0) \right| d\boldsymbol{v}\, d\boldsymbol{b} \right]^2 \quad \text{(A26)}$$

$$= \delta f_{\max}^2 \left[\sum_{i=1}^{n} \int_{\boldsymbol{R}^n} |v_i|\, |g(\boldsymbol{v})|\, d\boldsymbol{v} \int_{\boldsymbol{R}^n} |h(\boldsymbol{b})| \left| \sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0) \right| d\boldsymbol{b} \right]^2 . \quad \text{(A27)}$$

where $\delta f_{\max}$ is defined in (11), (12). Defining $\delta\sigma_{\max}$ in this way we can write

$$\left| \sigma(\boldsymbol{x}_0 - \boldsymbol{b}) - \sigma(\boldsymbol{x}_0) \right| \le \delta\sigma_{\max} \left| \sum_{j=1}^{n} b_j \right| \le \delta\sigma_{\max} \sum_{j=1}^{n} |b_j| . \quad \text{(A28)}$$

Then, using (A28) and by separating the sums and integrals the squared error (A23) is further bounded as:

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le \delta f_{\max}^2 \, \delta\sigma_{\max}^2 \Big[\sum_{i=1}^{n} \sum_{j=1}^{n} \int_{\boldsymbol{R}^n} |v_i| \, |g(\boldsymbol{v})| \, d\boldsymbol{v} \int_{\boldsymbol{R}^n} |b_i| \, |h(\boldsymbol{b})| \, d\boldsymbol{b} \Big]^2 \tag{A29}$$

$$= \delta f_{\max}^2 \, \delta\sigma_{\max}^2 \Big[\sum_{i=1}^{n} \int_{\boldsymbol{R}^n} |v_i| \, |g(\boldsymbol{v})| \, d\boldsymbol{v} \Big]^2 \Big[\sum_{j=1}^{n} \int_{\boldsymbol{R}^n} |b_i| \, |h(\boldsymbol{b})| \, d\boldsymbol{b} \Big]^2 . \tag{A30}$$

Using this special case of the Cauchy-Schwarz inequality:

$$\Big| \sum_{i=1}^{n} a_i \Big|^2 \le n \sum_{i=1}^{n} |a_i|^2 , \tag{A31}$$

then

$|\varepsilon(\boldsymbol{x}_0)|^2 \le$

$$n^2 \delta f_{\max}^2 \, \delta\sigma_{\max}^2 \Big[\sum_{i=1}^{n} \Big(\int_{\boldsymbol{R}^n} |v_i| \, |g(\boldsymbol{v})| \, d\boldsymbol{v} \Big)^2 \Big] \Big[\sum_{j=1}^{n} \Big(\int_{\boldsymbol{R}^n} |b_i| \, |h(\boldsymbol{b})| \, d\boldsymbol{b} \Big)^2 \Big] \tag{A32}$$

$$\le \mathrm{C_g} \, \mathrm{C_h} \, n^2 \delta f_{\max}^2 \, \delta\sigma_{\max}^2 \sum_{i=1}^{n} \int_{\boldsymbol{R}^n} v_i^2 |g(\boldsymbol{v})| d\boldsymbol{v} \cdot \sum_{j=1}^{n} \int_{\boldsymbol{R}^n} b_j^2 |h(\boldsymbol{b})| d\boldsymbol{b} \tag{A33}$$

where the second inequality (A33) follows since for a positive function $r(\boldsymbol{x})$ it is true that

$$\Big[\int_{\boldsymbol{R}^n} |x_i| r(\boldsymbol{x}) d\boldsymbol{x} \Big]^2 \le C_r \int_{\boldsymbol{R}^n} x_i^2 \, r(\boldsymbol{x}) \, d\boldsymbol{x} \qquad \text{where } C_r = \int_{\boldsymbol{R}^n} r(\boldsymbol{x}) \, d\boldsymbol{x} \tag{A34}$$

Therefore, we finally have

$$|\varepsilon(\boldsymbol{x}_0)|^2 \le n^2\, C_g\, C_h\, \delta f_{\max}^2\, \delta\sigma_{\max}^2\, Dg \cdot Dh \tag{A35}$$

which, upon taking square roots, completes the proof. ♣

Proof of Lemma 2.1 – With some simple substitutions we have that

$|\rho(\boldsymbol{x}, \boldsymbol{\xi})|^2 =$

$$\left| \int_{\boldsymbol{R}^n} g(\boldsymbol{a}) \int_{\boldsymbol{R}^n} g(\boldsymbol{b}) \left\{ R_f[\boldsymbol{\xi} - \boldsymbol{b}\sigma(\boldsymbol{x}+\boldsymbol{\xi}) + \boldsymbol{a}\sigma(\boldsymbol{x})] - R_f[\boldsymbol{\xi} - \boldsymbol{b}\sigma(\boldsymbol{x}) + \boldsymbol{a}\sigma(\boldsymbol{x})] \right\} d\boldsymbol{b}\, d\boldsymbol{a} \right|^2 \tag{A36}$$

$$\le \left[\int_{\boldsymbol{R}^n} |g(\boldsymbol{a})| \int_{\boldsymbol{R}^n} |g(\boldsymbol{b})| \left| R_f[\boldsymbol{\xi} - \boldsymbol{b}\sigma(\boldsymbol{x}+\boldsymbol{\xi}) + \boldsymbol{a}\sigma(\boldsymbol{x})] - R_f[\boldsymbol{\xi} - \boldsymbol{b}\sigma(\boldsymbol{x}) + \boldsymbol{a}\sigma(\boldsymbol{x})] \right| d\boldsymbol{b}\, d\boldsymbol{a} \right]^2 \tag{A37}$$

Reasoning as in (A24), (A25) we have that

$$|\rho(\boldsymbol{x}, \boldsymbol{\xi})|^2 \le \delta R_{f,\max}^2 \left[\int_{\boldsymbol{R}^n} |g(\boldsymbol{a})| \int_{\boldsymbol{R}^n} \left| \sum_{i=1}^n b_i \right| |g(\boldsymbol{b})| |\sigma(\boldsymbol{x}+\boldsymbol{\xi}) - \sigma(\boldsymbol{x})|\, d\boldsymbol{b}\, d\boldsymbol{a} \right]^2 \tag{A38}$$

$$\le C_g^2\, \delta R_{f,\max}^2\, \delta\sigma_{\max}^2 \left[\sum_{j=1}^n |\xi_j| \right]^2 \left[\int_{\boldsymbol{R}^n} \left| \sum_{i=1}^n b_i \right| |g(\boldsymbol{b})|\, d\boldsymbol{b} \right]^2 \tag{A39}$$

$$\le n^2 C_g^2\, \delta R_{f,\max}^2\, \delta\sigma_{\max}^2 \left[\sum_{j=1}^n |\xi_j| \right]^2 \left[\sum_{i=1}^n \int_{\boldsymbol{R}^n} |b_i|^2 |g(\boldsymbol{b})|\, d \right]^2 \tag{A40}$$

$$= n^2 C_g^2\, \delta R_{f,\max}^2\, \delta\sigma_{\max}^2\, (Dg)^2 \left[\sum_{j=1}^n |\xi_j| \right]^2 \tag{A41}$$

using (A28) and (A31). The proof is finished by taking the square root of both sides of (A41). ♣

Proof of Corollary 2.2 – The squared error can be written

$$|\varepsilon(\boldsymbol{m}_0)|^2 = \left| \sum_{\boldsymbol{p}\in\boldsymbol{Z}^n} h(\boldsymbol{p}) \left[\sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} f(\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}) g\left[\boldsymbol{r}/k(\boldsymbol{m}_0)\right] - \sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} f(\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}) g\left[\boldsymbol{r}/k(\boldsymbol{m}_0-\boldsymbol{p})\right] \right] \right|^2 \quad \text{(A41)}$$

$$= \left| \sum_{\boldsymbol{p}\in\boldsymbol{Z}^n} h(\boldsymbol{p}) \left[\sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} g(\boldsymbol{r}) f[\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}k(\boldsymbol{m}_0)] - f[\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}k(\boldsymbol{m}_0-\boldsymbol{p})] \right] \right|^2 . \quad \text{(A42)}$$

Reasoning similar to previous proofs, we have:

$$\left| f[\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}k(\boldsymbol{m}_0)] - f[\boldsymbol{m}_0-\boldsymbol{p}-\boldsymbol{r}k(\boldsymbol{m}_0-\boldsymbol{p})] \right| = \left| \sum_{i=1}^{n} \sum_{s_i=r_i k(\boldsymbol{m}_0)}^{r_i k(\boldsymbol{m}_0-\boldsymbol{p})} \nabla_i f(\boldsymbol{s}) \right| \quad \text{(A43)}$$

so that the squared error (A41) can be bounded

$$|\varepsilon(\boldsymbol{m}_0)|^2 \le \left[\sum_{i=1}^{n} \sum_{\boldsymbol{p}\in\boldsymbol{Z}^n} |h(\boldsymbol{p})| \left[\sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} |g(\boldsymbol{r})| \left| \sum_{s_i=r_i k(\boldsymbol{m}_0)}^{r_i k(\boldsymbol{m}_0-\boldsymbol{p})} \nabla_i f(\boldsymbol{s}) \right| \right] \right]^2 \quad \text{(A44)}$$

$$\le (\nabla f)^2_{\max} \left[\sum_{i=1}^{n} \sum_{\boldsymbol{p}\in\boldsymbol{Z}^n} |h(\boldsymbol{p})| \left[\sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} |r_i| |g(\boldsymbol{r})| |k(\boldsymbol{m}_0-\boldsymbol{p}) - k(\boldsymbol{m}_0)| \right] \right]^2 \quad \text{(A45)}$$

$$= (\nabla f)^2_{\max} \left[\sum_{i=1}^{n} \sum_{\boldsymbol{r}\in\boldsymbol{Z}^n} |r_i| |g(\boldsymbol{r})| \sum_{\boldsymbol{p}\in\boldsymbol{Z}^n} |h(\boldsymbol{p})| \, |k(\boldsymbol{m}_0-\boldsymbol{p}) - k(\boldsymbol{m}_0)| \right]^2 \quad \text{(A46)}$$

using (38) and (39). Now $|k(\boldsymbol{m}_0 - \boldsymbol{p}) - k(\boldsymbol{m}_0)| \le (\nabla k)_{\max} \sum_{j=1}^{n} |p_j|$ (A47)

where $(\nabla k)_{\max}$ is defined as in (38), (39). Hence (A46) is further bounded as

$$|\mathcal{E}(\boldsymbol{m}_0)|^2 \le (\nabla f)_{\max}^2 (\nabla k)_{\max}^2 \left[\sum_{i=1}^{n} \sum_{\boldsymbol{r} \in \boldsymbol{Z}^n} |r_i| |g(\boldsymbol{r})| \right]^2 \cdot \left[\sum_{j=1}^{n} \sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} |p_j| |h(\boldsymbol{p})| \right]^2 \quad \text{(A48)}$$

$$\le n^2 (\nabla f)_{\max}^2 (\nabla k)_{\max}^2 \sum_{i=1}^{n} \left[\sum_{\boldsymbol{r} \in \boldsymbol{Z}^n} |r_i| |g(\boldsymbol{r})| \right]^2 \cdot \sum_{j=1}^{n} \left[\sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} |p_j| |h(\boldsymbol{p})| \right]^2 \quad \text{(A49)}$$

$$\le n^2 (\nabla f)_{\max}^2 (\nabla k)_{\max}^2 \sum_{i=1}^{n} \sum_{\boldsymbol{r} \in \boldsymbol{Z}^n} r_i^2 |g(\boldsymbol{r})| \cdot \sum_{j=1}^{n} \sum_{\boldsymbol{p} \in \boldsymbol{Z}^n} p_j^2 |h(\boldsymbol{p})| \quad \text{(A50)}$$

$$\le n^2 (\nabla f)_{\max}^2 (\nabla k)_{\max}^2 \, Dg \cdot Dh \quad \text{(A51)}$$

using (A31) and the discrete version of (A34), (37) and (14). The proof is then finished by taking square roots. ♣

Appendix B

Definition, Proof and Algorithm of Chapter 3

B.1 Skewed Gaussian Probability Function

We define the skewed Gaussian to be a Gaussian with different standard deviations above (σ_h) and below (σ_l) the mode (u):

$$g(x;u,\sigma_l,\sigma_h)=\begin{cases}\dfrac{1}{\sqrt{2\pi}(\frac{\sigma_l+\sigma_h}{2})}\exp\left(\dfrac{-(x-u)^2}{2\sigma_l^2}\right) & if \quad x\le u\\ \dfrac{1}{\sqrt{2\pi}(\frac{\sigma_l+\sigma_h}{2})}\exp\left(\dfrac{-(x-u)^2}{2\sigma_h^2}\right) & if \quad c_a>u\end{cases} \tag{B1}$$

B.2 Differential Entropy of the Skewed Gaussian Distribution

The differential entropy of a probability density function is defined by the integral:

$$h(p)=\int_{-\infty}^{\infty}p(x)\ln[p(x)]dx \tag{B2}$$

Substituting Eq. (B1) into Eq. (B2) and letting $\phi_l(x)$ and $\phi_h(x)$ be Gaussian density functions with means u and standard deviations of σ_h and σ_l respectively, we have,

$$h(g)=\frac{2\sigma_l}{(\sigma_l+\sigma_h)}\int_{-\infty}^{u}\phi_l(x)\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_l}\right]+\ln\left(\sqrt{2\pi}\sigma_l\right)+\frac{(x-u)^2}{2\sigma_l^2}\right\}dx$$

$$+\frac{2\sigma_h}{(\sigma_l+\sigma_h)}\int_{-\infty}^{u}\phi_h(x)\left\{\ln\left[\frac{(\sigma_l+\sigma_h)}{2\sigma_h}\right]+\ln\left(\sqrt{2\pi}\sigma_h\right)+\frac{(x-u)^2}{2\sigma_h^2}\right\}dx$$

$$\Rightarrow h(g) = \frac{\sigma_l}{(\sigma_l + \sigma_h)}\left\{\ln\left[\frac{(\sigma_l + \sigma_h)}{2\sigma_l}\right] + \ln\left(\sqrt{2\pi\sigma_l^2}\right) + \frac{1}{2}\right\}$$

$$+ \frac{\sigma_h}{(\sigma_l + \sigma_h)}\left\{\ln\left[\frac{(\sigma_l + \sigma_h)}{2\sigma_h}\right] + \ln\left(\sqrt{2\pi\sigma_h^2}\right) + \frac{1}{2}\right\}$$

$$\Rightarrow h(g) = \frac{\sigma_l}{(\sigma_l + \sigma_h)}\left\{\ln\left[\frac{\sqrt{2\pi}(\sigma_l + \sigma_h)}{2}\right] + \frac{1}{2}\right\} + \frac{\sigma_h}{(\sigma_l + \sigma_h)}\left\{\ln\left[\frac{\sqrt{2\pi}(\sigma_l + \sigma_h)}{2}\right] + \frac{1}{2}\right\}$$

$$\Rightarrow h(g) = \ln\left[\frac{\sqrt{2\pi}(\sigma_l + \sigma_h)}{2}\right] + \frac{1}{2}$$

Thus we have that:

$$h(g) = \frac{1}{2}\log_2(2\pi e(\bar{\sigma})^2) \qquad \text{where, } \bar{\sigma} = \frac{\sigma_l + \sigma_h}{2}$$ ♣

B.3 Contrast Entropy Minimization Algorithm

Here we formalize the CEM algorithm. To begin with, let (x_i, y_i) represent the location of the *i*th pixel in the image, and let C_i be the true (unblurred) rms contrast at that location. We note that the term image location refers to a scene location expressed in degrees of visual angle in the horizontal and vertical directions.

Consider a series of fixations, *t*=1, 2,…. Let the location of fixation number *t* be x_t, y_t, and let the observed local rms contrast at the *i*th pixel, on that fixation, be c_{it}. The retinal eccentricity, ε_{it}, of the *i*th pixel location is

$$\varepsilon_{it} = \sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \qquad \text{(B3)}$$

Thus, if the observer is currently on fixation number *T*, then the current eccentricity map is given by

$$\varepsilon_{it}(T) = \min_{t \le T} \varepsilon_{it} \tag{B4}$$

(Note that new values appear in the eccentricity map only if a new fixation happens to bring a pixel closer to the fovea than it has been before.) The current contrast map, $c_i(T)$, is defined to be the contrast that was observed when the eccentricity was at its minimum value, as given by the eccentricity map [Eq. (B4)]. The uncertainty map is given by

$$h_i(T) = \frac{1}{2}\log_2\left(2\pi e\left\{\left[k\varepsilon_i(T)c_i(T)\right]^2 + \sigma_0^2\right\}\right) \tag{B5}$$

The total uncertainty after fixation number *T* is made is

$$U(T) = \sum_{i=1}^{n} h_i(T) \tag{B6}$$

To select the next fixation, the observer considers each possible location (x_{T+1}, y_{T+1}) for fixation *T+1*, estimates the total contrast uncertainty that will be obtained if that fixation is made, and then picks the location $(\hat{x}_{T+1}, \hat{y}_{T+1})$ with the minimum estimated total uncertainty:

$$(\hat{x}_{T+1}, \hat{y}_{T+1}) = \arg\max_{x_{T+1}, y_{T+1}} \left[\hat{U}(T+1, x_{T+1}, y_{T+1})\right] \tag{B7}$$

where,

$$\hat{U}(T+1, x_{T+1}, y_{T+1}) = \sum_{i=1}^{n} \hat{h}_i(T+1, x_{T+1}, y_{T+1}) \tag{B8}$$

$$\hat{h}_i(T+1, x_{T+1}, y_{T+1}) = \frac{1}{2}\log_2\left(2\pi e\left\{\left[k\varepsilon_i(T+1, x_{T+1}, y_{T+1})c_i(T+1, x_{T+1}, y_{T+1})\right]^2 + \sigma_0^2\right\}\right) \tag{B9}$$

To evaluate Eq. (B9), we note that the eccentricity map $\varepsilon_i(T+1, x_{T+1}, y_{T+1})$ for fixation location (x_{T+1}, y_{T+1}) is obtained directly from Eqs. (B3) and (B4). The estimated contrast

map, $\hat{c}_i(T+1, x_{T+1}, y_{T+1})$, can be obtained from text equation (3.1). Specifically, Eq. (3.1) gives the maximum *a posteriori* (MAP) estimate of the true contrast, $\hat{C}_i(T)$, for each location in the current contrast map:

$$\hat{C}_i(T) = kc_i(T)\varepsilon_i(T) + c_i(T) \tag{B10}$$

If this MAP estimate is relatively stable and unbiased, then approximately the same MAP estimate will be obtained after the next fixation is made,

$$\hat{C}_i(T) \cong kc_i(T+1, x_{T+1}, y_{T+1})\varepsilon_i(T+1, x_{T+1}, y_{T+1}) + c_i(T+1, x_{T+1}, y_{T+1}) \tag{B11}$$

and therefore our prediction of the observed contrast after the next fixation is

$$\hat{c}_i(T+1, x_{T+1}, y_{T+1}) = \frac{\hat{C}_i(T)}{k\varepsilon_i(T+1, x_{T+1}, y_{T+1}) + 1} \tag{B12}$$

In sum, Eq. (B3), (B4), (B7)–(B9), and (B12) can be used to estimate the fixation that will maximally reduce the total contrast uncertainty. In practice, we find that this estimate of the optimal fixation location is quite accurate.

A minor technical issue that arises in evaluating the CEM algorithm is that differential entropy can be negative. Therefore, we convert the differential entropy into discrete entropy by finely sampling the Gaussian distribution to obtain a discrete probability distribution. Using this discrete distribution guarantees that the uncertainty map is always nonnegative.

Appendix C

Proofs of Chapter 4

Proof of Theorem 4.1: We prove the lemma by induction on *d*. For the base case (*d*=2) it is easily shown that:

$$J(F) = \frac{1}{\beta_1 \beta_2 \,|\, C \,|} \prod_{k=1}^{2} \psi(\beta_k z_k) \cdot$$

Now assume by inductive hypothesis that the lemma is true for $d = 2, \ldots, N$. Consider the Jacobian when $d = N+1$:

$$J(F) = \begin{vmatrix} \frac{\partial F_1}{\partial s_1} & \cdot & \cdot & \frac{\partial F_1}{\partial s_{N+1}} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial F_{N+1}}{\partial s_1} & \cdot & \cdot & \frac{\partial F_{N+1}}{\partial s_{N+1}} \end{vmatrix} = \begin{vmatrix} a_{1,1}\psi(z_1) & \cdot & \cdot & a_{1,N+1}\psi(z_1) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{N+1,1}\psi(z_{N+1}) & \cdot & \cdot & a_{N+1,N+1}\psi(z_{N+1}) \end{vmatrix}$$

Expand $J(F)$ with respect to the first row:

$$J(F) = \left(\prod_{j=1}^{N+1} \beta_j \right)^{-1} \sum_{k=1}^{N+1} (-1)^{1+k} \,|\, C_{1,k} \,|\, a_{1,k} \psi(z_1)$$

where $C_{1,k}$ is the minor matrix of $J(F)$ with respect to $(1, k)$. Applying the inductive hypothesis yields:

$$J(F) = \frac{1}{\beta^{N+1}} \prod_{k=1}^{N+1} \psi(z_k) \sum_{k=1}^{N+1} (-1)^{1+k} \,|\, A_{1,k} \,|\, a_{1,k}$$

where $A_{1,k}$ is the minor matrix of A with respect to $(1, k)$. Thus,

$$J(F) = \frac{1}{\beta^{N+1}} \prod_{k=1}^{N+1} \psi(z_k) \,|\, A \,| = \frac{1}{\beta^{N+1} \,|\, C \,|} \prod_{k=1}^{N+1} \psi(z_k)$$

thereby proving the lemma for all *d*. ♣

Appendix D

Proofs of Chapter 5

Proof of Lemma 5.1: We have that

$$D\left[\mu^c(J_c)\,\|\prod_{k=1}^{d}\mu_m^c(I_k^c)\right] < D\left[\mu^s(J_s)\,\|\prod_{k=1}^{d}\mu_m^s(I_k^s)\right]$$

$$\Rightarrow E_{\mu^c}\left[\log\left(\frac{\mu^c(J_c)}{\prod_{k=1}^{d}\mu_m^c(I_k^c)}\right)\right] < E_{\mu^s}\left[\log\left(\frac{\mu^s(J_s)}{\prod_{k=1}^{d}\mu_m^s(I_k^s)}\right)\right]$$

$$\Rightarrow E_{\mu^c}\left[\log\left(\frac{\mu^c(J_c)}{\prod_{k=1}^{d}\mu_m^c(I_k^c)}\right)\right] < E_{\mu^s}\left[\log\left(\frac{\mu^s(J_s)}{\prod_{k=1}^{d}\mu_m^s(I_k^s)}\right)\right]$$

$$LHS = E_{\mu^c}\left[\log\left(g_c(J_c).\frac{\prod_{k=1}^{d}p_c(I_k^c)}{\prod_{k=1}^{d}\mu_m^c(I_k^c)}.\frac{\mu^c(J_c)}{g_c(J_c)\prod_{k=1}^{d}p_c(I_k^c)}\right)\right]$$

$$= \sum_{m\neq n} E_{\mu^c}\left[<\tilde{\varphi}(I_m^c),\tilde{\varphi}(I_n^c)>\right]G_{m,n}^c + C_1^c + C_2^c$$

where

$$C_1^c = D\left[\mu^c(J_c)\,\|\,g_c(J_c)\prod_{k=1}^{d}p_c(I_k^c)\right]$$

and $C_2^c = E_{\mu^c}\left[\ln\left(\prod_{k=1}^{d}\frac{p_c(I_k^c)}{\mu_m^c(I_k^c)}\right)\right]$.

A similar expression holds for the RHS. From this the lemma follows. ♣

Proof of Lemma 5.2: Since p is perfectly decomposable into its ICA components, then

$$D\left(p\,\|\prod_i p_i\right) = 0\,.$$

Therefore:

$$D\left(q \| \prod_i q_i\right) - D\left(p \| \prod_i p_i\right) = D\left(q \| \prod_i q_i\right)$$

$$= \int q \ln\left(\frac{q}{\prod_i q_i}\right) = \int q \ln\left(\frac{q}{p}\right) + \int q \ln\left(\frac{p}{\prod_i q_i}\right)$$

$$= D(q \parallel p) - \sum_i H[q_i] + E_q[\ln\ p]$$

Now since p is described by an MICA distribution, and given that it is perfectly decomposable into its ICA components, we have that:

$$p = \prod_i p_i$$

where

$$p(s_i) = K_i \exp\left(-\frac{1}{\sigma_i^2} \cdot \left\{\tilde{\varphi}\left[\beta_i\left(s_i - \mu_i\right)\right] - c_i\right\}^2\right)$$

Thus the lemma follows. ♣

Bibliography

[1] F. Attneave, "Some informational aspects of visual perception," *Psy. Review*, vol. 61, 1954, pp. 183–93.

[2] H.B. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory Communication*, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press, 1961.

[3] S.B. Laughlin, "A simple coding procedure enhances a neuron's information capacity," *Z. Naturforsch.* 36C:910–12, 1981.

[4] J.J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Netw. Comput. Neural Syst.* 3:213–51, 1992.

[5] J.H. van Hateren,"A theory of maximizing sensory information," *Biol. Cybern.* 68:23–29, 1992.

[6] D.J. Field, "What is the goal of sensory coding?," *Neural Comput.* 6:559–601, 1994.

[7] F. Rieke, D.A. Bodnar, W. Bialek, "Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents," *Proc. R. Soc. London B* 262:259–65, 1995.

[8] E.P. Simoncelli and O. Schwartz, Image statistics and cortical normalization models. In *Advances in Neural Information Processing Systems*, (MS Kearns, SA Solla, DA Cohn, Eds.), 11:153–59, 1999.

[9] B.A. Olshausen and D.J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* 381:607–9, 1996.

[10] A.J. Bell and T.J. Sejnowski, "The independent components" of natural scenes are edge filters," *Vis. Res.* vol. 37, no. 23, pp. 3327–38, 1997.

[11] T.D. Sanger, "Optimal unsupervised learning in a single-layer network," *Neural Netw.* 2:459–73, 1989.

[12] P. Foldiak, "Forming sparse representations by local anti-hebbian learning," *Biol. Cybernet.* 64:165–70, 1990

[13] J.H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc. R. Soc. London Ser. B* 265:359–66, 1998.

[14] B.A. Wandell, *Foundations of Vision,* Sinauer Associates Inc., 1995.

[15] A.C. Bovik and R.G. Raj, "Approximating filtered scale-variant signals," *IEEE Trans Image Process*, vol. 14, no. 1, pp. 23-35, Jan 2005.

[16] W. Geisler and J. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in *Human Vision and Electronic Imaging III*, vol. 3299. SPIE-Int. Soc. Opt. Eng, 1998, pp. 294–305.

[17] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley Interscience, 2001.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2001.

[19] W.S. Geisler, "Visual perception and the statistical properties of natural scenes," *Annual Review of Psychology*, 59, 10.1-10.26.

[20] S.B. Laughlin, "A simple coding procedure enhances a neuron's information capacity," *Z. Naturforsch.*, 36c: 910-2, 1981.

[21] D.L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems*, 5:517-548, 1994.

[22] N. Brady and D.J. Field, "Local contrast in natural images: normalization and coding efficiency," *Perception* 29: 1041-1055, 2000.

[23] D.L. Ruderman and W. Bialek, "Statistics of natural images: scaling in the woods," *Physics Review Letters*, 73: 814–817, 1994.

[24] Y. Tadmor and D.J. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes," *Vision Research* 40: 3145-3157, 2000.

[25] R.A. Frazor and W.S. Geisler, "Local luminance and contrast in natural images," *Vision Research*, 46:1585-1598, 2006.

[26] L.T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *Journal of the Optical Society of America A*, 3: 1673-83, 1986.

[27] L.T. Maloney and B.A. Wandell, "Color constancy: A method for recovering surface spectral reflectance," *Journal of the Optical Society of America A*, 3: 29-33, 1986.

[28] M. D'Zmura and G. Iverson, "Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces," *Journal of the Optical Society of America A*, 10:2148-2165, 1993.

[29] D.H. Brainard, and W.T. Freeman, "Bayesian color constancy," *Journal of the Optical Society of America A*, 14: 1393-411, 1997.

[30] D.J. Field, “Relations between the statistics of natural images and the repsonse properties of cortical cells,” *Journal of the Optical Society of America A*, 4: 2379-94, 1987.

[31] G.J. Burton and I.R. Moorehead, “Color and spatial structure in natural scenes,” *Applied Optics*, 26:157-170, 1987.

[32] M J Wainwright and E P Simoncelli, “Scale mixtures of Gaussians and the statistics of natural images,” in *Adv. Neural Information Processing Systems, S. A. Solla, T. K. Leen, and K.-R.Muller, Eds.*, Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press.

[33] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, “Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images,” *Applied and Computational Harmonic Analysis*, 11, 89-123, 2001.

[34] O. Schwartz, E.P. Simoncelli, “Natural signal statistics and sensory gain control,” *Nature Neuroscience*, 4: 819–825, 2001.

[35] J. Malo, I. Epifanio, R. Navarro, and E.P. Simoncelli, "Non-linear image representation for efficient perceptual coding," IEEE Trans Image Processing, 15(1):68-80, Jan 2006.

[36] A. Hyvarinen and P. Hoyer, “Emergence of topography and complex cell properties from natural images using extensions of ICA,” *In Advances in Neural Information Processing Systems, ed. SA Solla, TK Leen, K-R Muller*, pp. 827-33. Cambridge: MIT Press

[37] L. Yarbus, “Eye Movements and Vision,” *Kluwer Academic Publishers*, January 1967.

[38] L. E. Wixson and D. H. Ballard, “Using intermediate objects to improve the efficiency of visual search,” *Int. J. Computer Vision*, vol. 12:2-3, pp. 209–230, April 1994.

[39] T. Hertz, A. Bar-Hillel, and D. Weinshall, "Learning Distance Functions for Image Retrieval," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, Volume 2 pp. 570-577, 2004.

[40] A. Frome, Y. Singer, and J. Malik, "Image retrieval and classification using local distance functions," *In B. Scholkopf, J. Platt, and T. Hoffman, editors, Advances in Neural Information Processing Systems 19. MIT Press*, Cambridge, MA, 2007.

[41] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast Pose Estimation with Parameter-Sensitive Hashing," *Ninth IEEE International Conference on Computer Vision*, Volume 2, p.750-757, 2003.

[42] V. Athitsos and S. Sclaroff, "Boosting Nearest Neighbor Classi.ers for Multiclass Recognition," *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshops*, Volume 03, pp.45-52.

[43] J. Sivic, B. Russell, A.A. Efros, A. Zisserman, and B. Freeman, "Discovering Objects and Their Location in Images," *International Conference on Computer Vision (ICCV 2005)*, October, 2005.

[44] B.C. Russell, W.T. Freeman, A.A. Efros, J. Sivic, and A. Zisserman, "Using Multiple Segmentations to Discover Objects and their Extent in Image Collections," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 1605 - 1614, 2006.

[45] K. Grauman and T. Darrell, "Unsupervised Learning of Categories from Sets of Partially Matching Image Features," *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp.19 - 25, 2006.

[46] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning Object Categories from Google's Image Search," *Proceedings of the Tenth IEEE International Conference on Computer Vision*, Volume 2, pp. 1816 - 1823.

[47] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Volume 60, Issue 2, pp.91-110, November 2004.

[48] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," *Proceedings of the 7th European Conference on Computer Vision-Part I*, pp. 128 - 142, 2002.

[49] P. Moreels and P. Perona, "Evaluation of Features Detectors and Descriptors based on 3D Objects," *International Journal of Computer Vision*, Volume 73, Issue 3, pp.263-284, July 2007.

[50] F. Perronnin, C.R. Dance, G. Csurka, and M. Bressan, "Adapted Vocabularies for Generic Visual Categorization," *European Conference on Computer Vision*, vol. 4, pp.464-475, 2006.

[51] F. Moosmann, B. Triggs, F. Jurie, "Fast Discriminative Visual Codebooks using Randomized Clustering Forests," *Neural Information Processing Systems (NIPS)*, pp. 985-992, 2006.

[52] J. Winn, A. Criminisi, and T. Minka, "Object Categorization by Learned Universal Visual Dictionary," Proceedings of the Tenth IEEE International Conference on Computer Vision, Volume 2, pp. 1800 - 1807.

[53] K. Murphy, A. Torralba, and W.T. Freeman, "Using the Forest to See the Trees: A Graphical Model Relating Features, Objects, and Scenes," *Neural Information Processing Systems (NIPS)*, vol. 16, 2003.

[54] X. He, R.S. Zemel, and M.A. Carreira-Perpinan, "Multiscale Conditional Random Fields for Image Labeling," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 695-702, 2004.

[55] D. Hoiem, A.A. Efros, and M. Hebert, "Putting Objects in Perspective," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 2137-2144, 2006.

[56] X. Ren and J. Malik, "Learning a Classification Model for Segmentation," *Proceedings of the Ninth IEEE International Conference on Computer Vision*, Volume 2, pp.10-17, 2003.

[57] D.R. Martin, C.C. Fowlkes, and J. Malik, "Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 26 , Issue 5, pp. 530-549, May 2004.

[58] E. Borenstein and S. Ullman,"Class-Specific, Top-Down Segmentation," *Proceedings of the 7th European Conference on Computer Vision-Part II*, pp. 109 - 124, 2002.

[59] Z. Tu, X. Chen, A. Yuille, and S-C. Zhu, "Image parsing: Unifying segmentation, detection and recognition," *Int. J. Comput. Vis.*, vol. 63, pp. 113–140, 2005.

[60] A. Kannan, J. Winn, and C. Rother, "Clustering appearance and shape by learning jigsaws," *Neural Information Processing Systems (NIPS)*, 2006.

[61] W. S. Geisler and K.L. Chou, "Separation of low-level and high-level factors in complex tasks: Visual search," *Psy Review*, vol. 102, no. 2, pp. 356-378, 1995.

[62] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations," *IEEE Trans. Pattern Anal Machine Intell*, vol. 22, pp. 970–982, Sept 2000.

[63] G. J. Zelinsky, "Using eye saccades to assess the selectivity of search movements," *Vision Res*, vol. 36, pp. 2015–2228, July 1996.

[64] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Res*, vol. 40, no. 10-12, pp. 1489-1506, May 2000.

[65] B. Moghaddamand and A. Pentland, "Probabilistic visual learning for object detection," *Fifth Int'l Conf. Computer Vision*, pp. 786–793, June 1995.

[66] W. Klarquist and A. C. Bovik, "FOVEA: a foveated vergent active stereo system for dynamic three-dimensional scene recovery," *IEEE Tran. Robotics Automation*, vol. 14, pp. 755–770, Oct.1998.

[67] U. Rajashekar, I. van der Linde, A.C. Bovik, and L.K. Cormack, "GAFFE: A gaze-attentive fixation finding engine," *IEEE Trans Image Processing*, to appear.

[68] P. Reinagel and A. M. Zador, "Natural scene statistics at the center of gaze," *Network: Computation in Neural Systems*, vol. 10, no. 1-10, 1999.

[69] A.C. Bovik and R.G. Raj, "Approximating filtered scale-variant signals," *IEEE Trans Image Process*, vol. 14, no. 1, pp. 23-35, Jan 2005.

[70] R.G. Raj, W.S. Geisler, R.A. Frazor, and A.C. Bovik, "Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy, " *J. Opt Soc Amer A*, vol. 22, pp. 2039-2049, Oct 2005.

[71] R.G. Raj, W. S. Geisler, R. A. Frazor, and A. C. Bovik, "Contrast Statistics for Foveated Visual Systems: Contrast Constancy and Fixation Selection," *Vision Sciences Society 5th Annual Meeting, Sarasota, FL*, May 05-11, 2005.

[72] W.S. Geisler, R.A. Frazor, R.G. Raj, A.C. Bovik, V. Mante, and M. Carandini, "Local Luminance and Contrast in Natural Scenes: Implications for Understanding Visual Systems that Make Saccadic Eye Movements," *Sensory Coding And The Natural Environment*, Oxford, UK, September 05-10, 2004.

[73] R.G. Raj, W. S. Geisler, R. A. Frazor, and A. C. Bovik, "Natural Contrast Statistics and the Selection of Visual Fixations," *International Conference on Image Processing (ICIP) 2005*, Vol.3, 11-1, pp. 1152-5, September 2005.

[74] R.G. Raj and A.C. Bovik, "MICA: A Multilinear ICA Decomposition for Natural Image Modeling," *IEEE Trans Image Process*, August 2007, submitted.

[75] R.G. Raj and A.C. Bovik, "The Multilinear ICA Decomposition with Applications to NSS Modeling," *International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2007*, vol. 2, pp. 669-672, April 2007.

[76] R.G. Raj and A.C. Bovik, "Non-stationarity measurement in natural images," *IEEE Trans Image Process*, November 2007, submitted.

[77] R.G. Raj, A.C. Bovik, and W.S. Geisler, "Non-Stationarity Detection in Natural Images," *IEEE International Conference on Image Processing (ICIP 2007)*, Volume 3, Sept.16 2007-Oct.19 2007 Page(s):III - 305 - III - 308.

[78] R.G. Raj and A.C. Bovik, "Texture-Contrast based Fixation Selection in Natural Images," *Submitted to IEEE Transactions on Image Processing*

[79] S. Lee, M.S. Pattichis and A.C. Bovik, "Optimal rate control for real-time, low bitrate foveated video coding," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 977-992, July 2001.

[80] S. Lee, M.S. Pattichis and A.C. Bovik, "Foveated video quality assessment," *IEEE Transactions on Multimedia*, vol. 4, no. 1, pp. 129-132, March 2002.

[81] Z. Wang and A.C. Bovik, "Embedded foveation image coding," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1397-1410, October 2001.

[82] G.M. Robbins, "Image restoration for a class of linear spatially variant degradations," *Pattern Recognition*, vol. 2, pp. 91-103, 1970.

[83] A.A. Sawchuck, "Space-variant image restoration by coordinate transformation," *J. Opt. Soc. Am.*, vol. 64, pp. 138-144, Feb. 1974.

[84] M. Bolduc and M. D. Levine, "A review of biologically motivated space-variant data reduction models for robotic vision," *Computer Vision and Image Understanding*, vol. 69, pp. 170-184, 1998.

[85] C. Braccini, G. Gambardella, and G. Sandini, "A signal theory approach to the space and frequency variant filtering performed by the human visual system," *Signal Processing*, vol. 3, pp. 231-240, 1981.

[86] P. Perona, "Deformable kernels for early vision," *IEEE Trans. on Patt. Anal. And Mach. Intell.,* vol. 17, no. 5, pp. 488-499, 1995.

[87] P. Kortum and W. Geisler, "Implementation of a foveated image coding system for image bandwidth reduction," in *Proc. of the SPIE,* vol. 2657, pp. 350-360, San Jose, CA, Jan. 1996.

[88] W. T. Freeman and E. H. Adelson. "The design and use of steerable filters," *IEEE Trans. On Patt. Anal. and Mach. Intell.,* vol. 13, no. 9, pp. 891-906, 1991.

[89] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D.J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory,* vol. 38, no. 2, pp. 587-607, 1992.

[90] A. Tabernero, J. Portilla, and R. Navarro, "Duality of a log-polar image representation in the space and the spatial frequency domains," *IEEE Trans. on Sign. Proc.,* vol. 47, pp. 2469-79, Sep. 1999.

[91] A. Grossman and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape," *SIAM J. Math. Anal.,* vol. 15, no. 4, pp. 723 736, July 1984.

[92] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[93] B.K.P. Horn, *Robot Vision.* MIT Press, 1985.

[94] A.C. Bovik, P. Maragos and T.F. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE Transactions on Signal Processing,* Special Issue on Wavelets and Signal Processing, vol. 41, no. 12, pp. 3245-3265, December 1993.

[95] A. Papoulis, *The Fourier Integral and its Applications.* McGraw-Hill, 1962.

[96] A.C. Bovik (editor), *The Handbook of Image and Video Processing*, Academic Press, 2nd Edition, June 2005.

[97] S.O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282-332, July 1944; and vol. 24, pp. 46-156, Jan. 1945.

[98] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy by V1?" *Vision Res.*, **37**, 3311–3325 (1997).

[99] D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthalmic Physiol. Opt.*, 12, 229–232 (1992).

[100] J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?" *Neural Comput.*, 4, 196–210 (1992).

[101] J. H. van Hateren, "Real and optimal neural images in early vision," *Nature*, 360, 68–70 (1992).

[102] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly, "Edge co-occurrence in natural images predicts contour grouping performance," *Vision Res.* , 41, 711–724 (2001).

[103] D. Purves and R. B. Lotto, *Why We See What We Do: An Empirical Theory of Vision* (Sinauer, 2003).

[104] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, 24, 1193–1215 (2001).

[105] W. S. Geisler and R. Diehl, "Bayesian natural selection and the evolution of perceptual systems," *Philos. Trans. R. Soc. London, Ser. B*, 357, 419–448 (2002).

[106] P. L. Clatworthy, M. Chirimuuta, J. S. Lauritzen, and D. J. Tolhurst, "Coding of the contrasts in natural images by populations of neurons in primary visual cortex (VI)," *Vision Res.*, 43, 1983–2001 (2003).

[107] R. M. Balboa and N. M. Grzywacz, "Power spectra and distribution of contrasts of natural images from different habitats," *Vision Res.*, 43, 2527–2537 (2003).

[108] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," *IEEE Trans. Pattern Anal. Mach. Intell.*, 18, 1–14 (1996).

[109] T. S. Lee and S. Yu, "An information-theoretic framework for understanding saccadic behaviors," in *Advances in Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Muller, eds. (MIT Press, 2000) Vol. 12, pp.834–840.

[110] G. E. Legge, T. A. Hooven, T. S. Klitz, J. G. Mansfield, and B. S. Tjan, "Mr. Chips 2002: new insights from an ideal observer model of reading," *Vision Res.*, 42, 2219–2234 (2002).

[111] L. W. Renninger, J. Coughlan, P. Verghese, and J. Malik, "An information maximization model of eye movements," in *Advances in Neural Information Processing Systems*, 17, L. K. Saul, Y. Weiss, and L. Bottou, eds. (MIT Press, 2005), pp. 1121–1128.

[112] J. G. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vision Res.*, 21, 409–418 (1981).

[113] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," *J. Opt. Soc. Am. A*, 8, 1775–1787 (1991).

[114] T. L. Arnow and W. S. Geisler, "Visual detection following retinal damage: predictions of an inhomogeneous retinocortical model," *Proc. SPIE*, 2674, 119–130 (1996).

[115] D. C. Hood and M. A. Finkelstein, "Sensitivity to light," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, 1986), Vol. 1.

[116] T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, 1991).

[117] U. Rajashekar, L. K. Cormack, and A. C. Bovik, "Visual search: structure from noise," in *Proceedings of Eye Tracking Research & Applications, ACM SIGGRAPH* 2002, A. T. Duchowski, ed. pp. 119–123.

[118] J. Najemnik and W. S. Geisler, "Optimal eye movement strategies in visual search," *Nature*, 434, 387–391 (2005).

[119] J-F. Cardoso, "Blind source separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[120] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[121] A. Hyvarinen, P.O. Hoyer, and M. Inki, "Topographic independent component analysis," *Neural Comput.*, vol. 13, no. 7, pp. 1527–1558, July 2001.

[122] N. Balakrishnan and C.R. Rao (editors), *Handbook of Statistics 16: Order Statistics: Theory & Methods, Elsevier*, 1998.

[123] USC Signal & Image Processing Institute Image Database, Volume 1: Textures. [online]. Available at: http://sipi.usc.edu/database/database.cgi?volume=textures.

[124] P. Comon, “Independent component analysis: A new concept?,” *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[125] U. Rajashekar, I. van der Linde, A.C. Bovik and L.K. Cormack, “Foveated analysis of image features at fixations,” *Vision Res.*, in press.

[126] S.C. Zhu, Y. Wu, and D. Mumford, “Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling,” *Int’l J Comput. Vision*, vol.27, no.2, pp.107-126, March 1998.

[127] R. Chellappa and A. Jain (editors), *Markov Random Fields: Theory and Application*, Academic Press, April 1993.

[128] J. Portilla and E.P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int’l J Comput. Vision*, vol. 40, (1), pp. 49-70, October 2000.

[129] C.H. Chen, L.F. Pau, and P.S.P. Wang, *Handbook of Pattern Recognition and Computer Vision*, World Scientific Publishing Company, 2nd Edition, April 1999.

[130] D. Donoho, S. Mallat, and R. von Sachs, "Estimating covariances of locally stationary processes: Consistency of best basis methods," *Proc. Time Frequency and Time-Scale Symposium*, Paris, July 1996.

[131] S. Mallat, Z. Zhang, and G. Papanicolaou, "Adaptive covariance estimation of locally stationary processes," *Ann. Stat.*, vol.26, no.1, pp.1-47, 1998.

[132] R. Dahlhaus, "On the Kullback-Leibler information divergence of locally stationary processes," *Stochastic Processes Applications*, 1995.

[133] M.B. Priestley, "Evolutionary spectra and non-stationary processes," *J Royal Soc. Statist.*, B, vol.27, pp. 204-229, 1965.

[134] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of non-stationary processes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol.33, no.6, pp. 1461-1470, December 1985.

[135] G. Matz, F. Hlawatsch, and W. Kozek, "Generalized evolutionary spectral analysis and the Weyl spectrum of non-stationary random processes," *Technical Report, 95-04*, Department of Communications, Vienna University of Technology, 1995.

[136] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "Multidimensional Quasi-eigenfunction Approximations and Multicomponent AM-FM Models," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 227- 242, Feb 2000.

[137] S. Bochner, "Stationarity, boundedness, and almost-periodicity of random valued functions," *Proc. Third Berkeley Symp. Math. Statist. and Prob.* 2 (1956), pp. 7-27.

[138] K. Karhunen, "Uber lineare methoden in der wahrscheinlichkeitsrechung," *Ann. Acad. Sci. Fennicae*, Ser. A, 1947, pp. 3-47.

[139] M. Loeve, *Probability Theory*, (3rd edition), D. van Nostrand, Princeton, NJ, 1963.

[140] H. Cramer, "A contribution to the theory of stochastic processes," *Proc. Second Berkeley Symp. Math.. Statist. and Prob.*, 1951, 329-339.

[141] M.M. Rao, "Harmonizable processes: Structure theory," *L'Enseign Math.* 28 (1982), 295-351.

[142] L. L. Scharf, P. J. Schreier, and A. Hanssen, "The Hilbert space geometry of the Rihaczek distribution for stochastic analytic signals," *IEEE Signal Process. Letters*, vol. 12, no. 4, pp. 297–300, Apr. 2005.

[143] B. Pesquet-Poescu and J.L. Vehel, "Stochastic fractal models for image processing," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 48-62, September 2002.

[144] E. Carlstein, H-G. Muller, and D. Siegmund (Editors), "*Change-point Problems*," Lecture Notes – Monograph Series, Institute of Mathematical Statistics, Feb 1994.

[145] A. Hyavarinen, Erkki Oja, Patrik Hoyer and Jarmo Hurri, "Image feature extraction by sparse coding and independent component analysis," *Proc 14th Int'l Conf Pattern Recogn*, vol. 2, pp. 1268-1273, 1998.

[146] U. Rajashekar, I. van der Linde, A. C. Bovik, and L. K. Cormack, "Foveated analysis and selection of visual fixations in natural scenes," *IEEE Int'l Conf Image Process.*, Oct 12-15, 2006.

[147] Z. Wang, L. Lu and A.C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 243-254, Feb 2003.

[148] A.C. Bovik, M. Clark, and W.S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans Pattern Anal Machine Intell*, vol. 12, no. 1, January 1990, pp. 55-73.

[149] A.C. Bovik, "Analysis of multichannel narrow-band filters for image texture segmentation," *IEEE Trans Signal Process*, vol. 39, no. 9, pp. 2025-2043 Sept 1991.

[150] A.C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo (Palacios), "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE Trans Info Theory*, vol. 38, no. 2, March 1992, pp. 2025-2043.

[151] I. van der Linde, U. Rajashekar, A. C. Bovik, L. K. Cormack, "DOVES: A database of visual eye movements," *Spatial Vision*, submitted, 2007.

[152] I. van der Linde, U. Rajashekhar, A.C. Bovik and L.K. Cormack, *DOVES: A Database of Visual Eye Movements,* July 2007 [Online]. Available: http://live.ece.utexas.edu/research/doves.

[153] A. Papoulis and S.U. Pillai, *Probability, Random Variables and Stochastic Processes,* 4th Edition, McGraw Hill, 2002.

[154] S. Kullback and R.A. Leibler, "On information and sufficiency," *Ann Math Stat*, vol. 22, pp. 79-86, 1951.

[155] D. Johnson and S. Sinanovic, "Symmetrizing the Kullback-Leibler distance," unpublished manuscript available online: http://www.ece.rice.edu/~dhj/cv.html.

[156] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal Machine Intel*, vol. 20, no. 11, pp. 1254-1259, 1998.

[157] S.L. Lauritzen, *Graphial Models*, Oxford University Press, Oxford, 1996.

[158] S. Amari, "Information Geometry on Hierarchy of Probability Distributions," *IEEE Transactions on Information Theory*, Vol. 47, No. 5, July 2001.

Vita

Raghu G. Raj was born in Augusta, GA, USA on June 21, 1975. He received his B.S. degrees in Computer Science and Electrical Engineering from Washington University, St. Louis, MO, in 1998 during which period he also worked for two semesters as a co-op engineer in MagneTek Inc., St. Louis. Thereafter he received his M.S. degree in Electrical Engineering from the University of Texas at Austin in May 2000 during which period he was a Graduate Research Assistant in the Advanced Sonar Division of the Applied Research Laboratories (ARL) in UT-Austin.

Thereafter he was with Motorola, Inc., Austin, TX, for four years, working on signal processing applications for communication systems. He then pursued the Ph.D. degree in Electrical Engineering in the area of Visual Search. His research interests include visual search, automatic target recognition, signal/image processing, computational vision, multi-dimensional stochastic processes, pattern recognition, data mining, and data compression.

Permanent Address: 3517 North Hills Drive, Apt. H201,

Austin, TX 78731

This dissertation was typed by the author